DOCUMENT RESUME

ED 480 803 TM 035 224

AUTHOR	Thurgood, Lori; Walter, Elizabeth; Carter, George; Henn, Susan; Huang, Gary; Nooter, Daniel; Smith, Wray; Cash, R.William; Salvucci, Sameena; Seastrom, Marilyn; Phan, Tai; Cohen, Michael
TITLE	NCES Handbook of Survey Methods: Technical Report.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.; Synectics for Management Decisions, Inc., Arlington, VA.
REPORT NO	NCES-2003-603
PUB DATE	2003-04-00
NOTE	339p.
CONTRACT	RN96020001
AVAILABLE FROM	EDPubs, P.O. Box 1398, Jessup, MD 20794-1398. Tel: 877-433-7827 (Toll Free).
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	EDRS Price MF01/PC14 Plus Postage.
DESCRIPTORS	Data Analysis; *Data Collection; Evaluation Utilization; Research Design; *Research Methodology; *Surveys

ABSTRACT

This handbook presents current explanations of how each survey program of the National Center for Education Statistics (NCES) obtains and prepared the data it publishes. The handbook aims to provide users of NCES data with the most current information necessary to evaluate the suitability of the statistics for their needs, with a focus on the methodologies for survey design, data collection, and data processing. The handbook contains 18 chapters. Chapters 1 to 26 each focus on one of the 26 major NCES survey programs. Information in these chapters is presented in a uniform format of: (1) overview; (2) uses of data; (3) key concepts; (4) survey design; (5) data quality and comparability; (6) contact information; and (7) methodology and evaluation reports. These chapters are organized into these subject matter categories: (1) early childhood education survey; (2) elementary and secondary education surveys; (3) library surveys; (4) postsecondary and adult education surveys; (5) educational assessment surveys; and (6) household education surveys. The final chapters, 27 and 28, cover multiple surveys and survey systems, including the Fast Response Survey System. Five appendixes contain a glossary, information on how to obtain NCES publications, a list of tools for each survey, a list of Web sites for each survey, and an index. (SLD)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 480 803

NCES
National Center for Education Statistics

U.S. Department of Education
Institute of Education Sciences
NCES 2003-603

NCES Handbook of Survey Methods

Technical Report

April 2003

Lori Thurgood
Elizabeth Walter
George Carter
Susan Henn
Gary Huang
Daniel Nooter
Wray Smith
R. William Cash
Sameena Salvucci
Synectics for Management Decisions, Inc.

Marilyn Seastrom, Tai Phan, and Michael Cohen, Project Officers
National Center for Education Statistics

TM035224

BEST COPY AVAILABLE

Acknowledgments

This first edition of the *NCES Handbook of Survey Methods* would not have been possible without the efforts of many people. First, we acknowledge the expertise of the NCES Oversight Committee, led by Marilyn Seastrom. The members of the Committee reviewed all chapters and provided insightful comments on both the format and accuracy of the survey descriptions. Members of the Oversight Committee are: Jonaki Bose, Michael Cohen, Dan Kasprzyk, Steven Kaufman, Andrew Kolstad, Bruce Taylor, and Tai Phan.

Sincere thanks to all the NCES project officers on the surveys and studies documented in the Handbook—who answered questions, provided unpublished materials, and reviewed the chapters about their surveys. Their names are given as project contacts in the chapters relevant to their surveys. Some of these individuals also served on the Oversight Committee and provided comments for the adjudication.

We express appreciation to Tracy Ferbish and Aubrey Kulms, of the National Education Data Resource Center at Pinkerton Computer Consultants, for their constant good cheer in locating older publications and identifying recent releases of reports and data files; Libby Farris, of Westat, for providing documentation on the early years of the Fast Response Survey; Jennifer Day, of the U.S. Bureau of the Census, for providing much-needed documentation on the Current Population Survey and the October Supplement; Susan Hill, of the National Science Foundation, for reviewing the chapter on the Survey of Earned Doctorates; and Evelyn Powers, of Design Powers, for designing the desktopped format for the final publication.

Karen O'Conor, Mathematical Statistician, Statistical Standards Program, chaired the adjudication committee. Individuals within and outside the Department of Education provided review comments for the adjudication: from NCES: Stephen Broughman, Shelley Burns, Kathryn Chandler, Chris Chapman, Steven Gorman, William Hussar, Steven Kaufman, Paula Knepper, Ralph Lee, Dawn Nelson, Jerry West, and Jeff Williams; from the Office for Civil Rights, Mary Schifferli; from the Office of Vocational and Adult Education, Ricardo Hernandez; from the Office of Management and Budget, Brian Harris-Kojetin; from the Bureau of Labor Statistics, Steven Cohen; and Education Statistics Services Institute (ESSI) staff who work on the *Education Statistics Quarterly*.

This Handbook was prepared by Synectics for Management Decisions, Inc., under contract No. RN96020001. Marilyn Seastrom, Tai Phan, and Michael Cohen served as NCES project officers. Lori Thurgood led the Synectics team, under the direction of Sameena Salvucci, Director of Research and Statistical Services. The authors wish to recognize the substantial contributions of several other Synectics staff: for assistance in developing, writing, and reviewing chapters, Binbing Yu and Fan Zhang; and for reviewing statistical sections of chapters, Dhiren Ghosh and Maxime Bokossa.

BEST COPY AVAILABLE

Contents

Acknowledgments ... iii

List of Acronyms ... xi

Introduction ... 1
NCES Role and Organization ... 1
Organization of the Handbook ... 1

EARLY CHILDHOOD EDUCATION SURVEY

Chapter 1: Early Childhood Longitudinal Study (ECLS) ... 5
1. Overview ... 5
2. Uses of Data ... 7
3. Key Concepts ... 7
4. Survey Design ... 8
5. Data Quality and Comparability ... 15
6. Contact Information ... 17
7. Methodology and Evaluation Reports ... 17

ELEMENTARY AND SECONDARY EDUCATION SURVEYS

Chapter 2: Common Core of Data (CCD) ... 19
1. Overview ... 19
2. Uses of Data ... 20
3. Key Concepts ... 21
4. Survey Design ... 21
5. Data Quality and Comparability ... 23
6. Contact Information ... 25
7. Methodology and Evaluation Reports ... 25

Chapter 3: Private School Universe Survey (PSS) ... 27
1. Overview ... 27
2. Uses of Data ... 27
3. Key Concepts ... 28
4. Survey Design ... 28
5. Data Quality and Comparability ... 31
6. Contact Information ... 33
7. Methodology and Evaluation Reports ... 33

Chapter 4: Schools and Staffing Survey (SASS) ... 35
1. Overview ... 35
2. Uses of Data ... 37
3. Key Concepts ... 38
4. Survey Design ... 38
5. Data Quality and Comparability ... 43
6. Contact Information ... 45
7. Methodology and Evaluation Reports ... 45

Chapter 5: SASS Teacher Follow-up Survey (TFS) 47
1. Overview 47
2. Uses of Data 48
3. Key Concepts 48
4. Survey Design 48
5. Data Quality and Comparability 50
6. Contact Information 51
7. Methodology and Evaluation Reports 51

Chapter 6: National Education Longitudinal Study of 1988 (NELS:88) 53
1. Overview 53
2. Uses of Data 55
3. Key Concepts 56
4. Survey Design 57
5. Data Quality and Comparability 61
6. Contact Information 65
7. Methodology and Evaluation Reports 65

Chapter 7: National Longitudinal Study of the High School Class of 1972 (NLS-72) 67
1. Overview 67
2. Uses of Data 69
3. Key Concepts 69
4. Survey Design 70
5. Data Quality and Comparability 75
6. Contact Information 78
7. Methodology and Evaluation Reports 78

Chapter 8: High School and Beyond (HS&B) Longitudinal Study 81
1. Overview 81
2. Uses of Data 83
3. Key Concepts 84
4. Survey Design 84
5. Data Quality and Comparability 88
6. Contact Information 91
7. Methodology and Evaluation Reports 91

LIBRARY SURVEYS

Chapter 9: SASS School Library Survey (SLS) 93
1. Overview 93
2. Uses of Data 94
3. Key Concepts 94
4. Survey Design 94
5. Data Quality and Comparability 96
6. Contact Information 96
7. Methodology and Evaluation Reports 97

Chapter 10: Public Libraries Survey (PLS) 99
1. Overview 99
2. Uses of Data 100
3. Key Concepts 100
4. Survey Design 101
5. Data Quality and Comparability 102

6. Contact Information 103
7. Methodology and Evaluation Reports 103

Chapter 11: Academic Libraries Survey (ALS) 105
1. Overview 105
2. Uses of Data 106
3. Key Concepts 106
4. Survey Design 107
5. Data Quality and Comparability 110
6. Contact Information 110
7. Methodology and Evaluation Reports 110

Chapter 12: State Library Agencies (StLA) Survey 111
1. Overview 111
2. Uses of Data 111
3. Key Concepts 111
4. Survey Design 112
5. Data Quality and Comparability 114
6. Contact Information 114
7. Methodology and Evaluation Reports 114

Chapter 13: Federal Libraries and Information Centers Survey 115
1. Overview 115
2. Uses of Data 115
3. Key Concepts 116
4. Survey Design 117
5. Data Quality and Comparability 118
6. Contact Information 118
7. Methodology and Evaluation Reports 119

POSTSECONDARY AND ADULT EDUCATION SURVEYS

Chapter 14: Integrated Postsecondary Education Data System (IPEDS) 121
1. Overview 121
2. Uses of Data 125
3. Key Concepts 126
4. Survey Design 127
5. Data Quality and Comparability 131
6. Contact Information 134
7. Methodology and Evaluation Reports 134

Chapter 15: National Study of Postsecondary Faculty (NSOPF) 135
1. Overview 135
2. Uses of Data 136
3. Key Concepts 137
4. Survey Design 138
5. Data Quality and Comparability 143
6. Contact Information 147
7. Methodology and Evaluation Reports 147

Chapter 16: National Postsecondary Student Aid Study (NPSAS) 149
1. Overview 149
2. Uses of Data 150
3. Key Concepts 150

4. Survey Design 151
5. Data Quality and Comparability 156
6. Contact Information 158
7. Methodology and Evaluation Reports 158

Chapter 17: Beginning Postsecondary Students (BPS) Longitudinal Study 161
1. Overview 161
2. Uses of Data 162
3. Key Concepts 162
4. Survey Design 163
5. Data Quality and Comparability 165
6. Contact Information 167
7. Methodology and Evaluation Reports 167

Chapter 18: Baccalaureate and Beyond (B&B) Longitudinal Study 169
1. Overview 169
2. Uses of Data 170
3. Key Concepts 171
4. Survey Design 171
5. Data Quality and Comparability 175
6. Contact Information 177
7. Methodology and Evaluation Reports 177

Chapter 19: Survey of Earned Doctorates (SED) 179
1. Overview 179
2. Uses of Data 180
3. Key Concepts 180
4. Survey Design 181
5. Data Quality and Comparability 183
6. Contact Information 186
7. Methodology and Evaluation Reports 186

EDUCATIONAL ASSESSMENT SURVEYS

Chapter 20: National Assessment of Educational Progress (NAEP) 187
1. Overview 187
2. Uses of Data 190
3. Key Concepts 191
4. Survey Design 191
5. Data Quality and Comparability 200
6. Contact Information 203
7. Methodology and Evaluation Reports 204

Chapter 21: Third International Mathematics and Science Study (TIMSS) 207
1. Overview 207
2. Uses of Data 210
3. Key Concepts 211
4. Survey Design 211
5. Data Quality and Comparability 219
6. Contact Information 221
7. Methodology and Evaluation Reports 221

Chapter 22: IEA Reading Literacy Study 223
1. Overview 223
2. Uses of Data 224
3. Key Concepts 224
4. Survey Design 224
5. Data Quality and Comparability 229
6. Contact Information 230
7. Methodology and Evaluation Reports 230

Chapter 23: National Adult Literacy Survey (NALS) 231
1. Overview 231
2. Uses of Data 232
3. Key Concepts 233
4. Survey Design 233
5. Data Quality and Comparability 239
6. Contact Information 241
7. Methodology and Evaluation Reports 241

Chapter 24: International Adult Literacy Survey (IALS) 243
1. Overview 243
2. Uses of Data 244
3. Key Concepts 244
4. Survey Design 245
5. Data Quality and Comparability 250
6. Contact Information 253
7. Methodology and Evaluation Reports 253

HOUSEHOLD SURVEYS

Chapter 25: National Household Education Surveys (NHES) Program 255
1. Overview 255
2. Uses of Data 257
3. Key Concepts 258
4. Survey Design 258
5. Data Quality and Comparability 264
6. Contact Information 268
7. Methodology and Evaluation Reports 269

Chapter 26: Current Population Survey (CPS)—October and September Supplements 271
1. Overview 271
2. Uses of Data 272
3. Key Concepts 272
4. Survey Design 273
5. Data Quality and Comparability 275
6. Contact Information 276
7. Methodology and Evaluation Reports 277

SMALL SPECIAL-PURPOSE NCES SURVEYS

Chapter 27: Fast Response Surveys 279
1. Fast Response Survey System (FRSS) 279
2. Postsecondary Education Quick Information System (PEQIS) 283

Chapter 28: Other NCES Surveys and Studies ... 287
1. School Crime Supplement (SCS) ... 287
2. School Survey on Crime and Safety (SSOCS) ... 291
3. High School Transcript (HST) Studies ... 292
4. Library Cooperatives Survey (LCS) ... 300
5. Civic Education Study (CivEd) ... 301

APPENDICES

Appendix A: Glossary of Statistical Terms ... A-1
Appendix B: Ordering NCES Publications and Data Files ... B-1
Appendix C: Web-based and Standalone Tools for Use with NCES Survey Data ... C-1
Appendix D: NCES Survey Web Site Addresses ... D-1
Appendix E: Index ... E-1

List of Acronyms

3PL	Three-parameter logistic scaling model

AE	Adult Education Survey (see NHES)
AELL	Adult Education and Lifelong Learning Survey (see NHES)
AERA	American Educational Research Association
ALA	American Library Association
ALA-ORS	American Library Association's Office of Research and Statistics
ALL	Adult Literacy and Lifeskills Survey (see NALS)
ALS	Academic Library Survey
API	Asian/Pacific Islander
APGAR	Activity (muscle tone)/Pulse/Grimace (reflex irritability)/Appearance (skin color)/Respiration; the APGAR score is given to newborns (see ECLS)
AQS	Attachment Q-Sort (see ECLS)
ASPA	Before- and After-School Programs and Activities Survey (see NHES)

B&B	Baccalaureate and Beyond Longitudinal Study
BHR	Balanced Half-sample Replication
BHS	Balanced Half-sample (BHS) method (weighing)
BIA	Bureau of Indian Affairs
BIB	Balanced Incomplete Block Spiraling (a type of matrix sampling)
BILOG	computer software used in scaling (see IALS)
BJS	Bureau of Justice Statistics, U.S. Department of Justice
BLS	Bureau of Labor Statistics, U.S. Department of Labor
BPS	Beginning Postsecondary Students Longitudinal Study
BRR	Balanced Repeated Replication
BSID	Bayley Scales for Infant Development (see ECLS)
BYI	Base Year Ineligible Study (see NELS:88)

C	Completions (see IPEDS)
CACE	Computer Assisted Coding and Editing (see HST)
CADE	Computer-Assisted Data Entry
CAO	Chief Administrative Officer (NSOPF)
CAPI	Computer-Assisted Personal Interview
CASES	Computer-Assisted Survey Execution System (see SCS)
CATI	Computer-Assisted Telephone Interviewing
CCD	Common Core of Data
CCSSO	Council of Chief State School Officers
CEO	Chief Executive Officer
CES	Civic Education Study (see CivEd)
CHAID	Chi-Squared Automatic Interaction Detector
CI	Civic Involvement Survey (see NHES)
CIDS	Career Information Delivery Systems (see IPEDS)
CIP	Classification of Instructional Programs
CivEd	Civic Education Study
CN	Consolidated Form (see IPEDS)
CN-F	Consolidated Form, "finance" part (see IPEDS)
COSLA	Chief Officers of State Library Agencies

CPS	Current Population Survey
CSSC	Classification of Secondary School Courses (see HST)
DAS	Data Analysis System
DIF	Differential Item Functioning (see NAEP)
DOD	Department of Defense
DODEA	Department of Defense Education Activity
DOL	Department of Labor
DRC	Data Recognition Corporation
DRF	Doctorates Record File (see SED)
EA	Institutional Activity survey (see IPEDS)
EAP	Employees by Assigned Position (see IPEDS)
ECB	Electronic Code Book
ECE	Early Childhood Education Survey (see NCES)
ECLS	Early Childhood Longitudinal Study
ECLS-B	Early Childhood Longitudinal Study, Birth Cohort
ECLS-K	Early Childhood Longitudinal Study, Kindergarten Cohort
ECPP	Early Childhood Program Participation Survey (see NHES)
ED	United States Department of Education
EDI	Electronic Data Interchange (see NPSAS)
EEOC	U.S. Equal Employment Opportunity Commission (see IPEDS)
EF	Fall Enrollment (see IPEDS)
EFC	Expected Family Contribution (see NPSAS)
EIAC	Education Information Advisory Council
EP	Fall Enrollment in Occupationally-specific Programs (see IPEDS)
ESL	English as a Second Language
ETS	Educational Testing Service
F	Finance (see IPEDS)
FASB	Financial Accounting Standards Board (see IPEDS)
FEDLINK	Federal Library and Information Network (see Federal Libraries)
FICE	Federal Interagency Committee on Education
FIMS	First International Mathematics Study (see TIMSS)
FIPS	Federal Information Processing Standards
FISS	First International Science Study (see TIMSS)
FLICC	Federal Library and Information Center Committee (see Federal Libraries)
FMS	Field Management System (see ECLS)
FRSS	Fast Response Survey System
FSCS	Federal-State Cooperative System (see PLS)
FSES	Followback Study of Excluded Students (see NELS:88)
FTB	First-time beginning/beginners (see BPS)
FTE	Full-time Equivalency
FY	Fiscal Year
GASB	Government Accounting Standards Board
GED	General Equivalency Diploma
GPA	Grade-Point Average
GPC	Generalized Partial Credit model (see CivEd and NAEP)
GRS	Graduation Rate Survey (see IPEDS)
GVF	Generalized Variance Function

HEGIS	Higher Education General Information System
HHL	Household and Library Use Survey (see NHES)
HLM	Hierarchical Linear Modeling
HS&B	High School and Beyond
HSES	High School Effectiveness Study (see NELS:88)
HST	High School Transcript studies
IAEP	International Assessment of Educational Progress (see NAEP)
IALS	International Adult Literacy Survey
IC	Institutional Characteristics (see IPEDS)
ICC	International Coordinating Center
ICS	Integrated Control System (see BPS)
IDEALS	microcomputer program for reporting library data
IEA	International Association for the Evaluation of Educational Achievement
IEP	Individual Education Program (see CCD and IEA Reading) or Individualized Educational Program (see HST) or Individualized Education Plan (see NAEP)
IES	Institute of Education Sciences
IHE	Institution of Higher Education
IPEDS	Integrated Postsecondary Education Data System
IRT	Item Response Theory
ISA	International Study of Adults (see IALS)
ISC	International Steering Committee
ISCED	International Standard Classification of Education (see IALS)
ISIC	International Standard Industrial Classification (see IALS)
ISOC	International Standard Occupational Classification (see IALS)
JRR	Jackknife Repeated Replication
LEA	Local Education Agency
LCS	Library Cooperatives Survey
LEP	Limited English Proficiency
LSCA	Library Services and Construction Act (see StLA)
LSTA	Library Services and Technology Act
MAS	Massey Attachment Sort (see ECLS)
MIL	Missing Information Letter (see SED)
MML	Marginal Maximum Likelihood (see NAEP)
MOSS	Method of Successive Sorts (see ECLS)
MSA	Metropolitan Statistical Area
NAAL	National Assessments of Adult Literacy
NAEP	National Assessment of Educational Progress
NAGB	National Assessment Governing Board (see NAEP)
NALS	National Adult Literacy Survey
NCATS	Nursing Child Assessment Teaching Scale (see ECLS)
NCEA	National Catholic Educational Association
NCES	National Center for Education Statistics
NCHS	National Center for Health Statistics
NCLIS	U.S. National Commission on Libraries and Information Science
NCS	National Computer Systems
NCVS	National Crime Victimization Survey

NELS:88	National Education Longitudinal Study of 1988
NHES	National Household Education Survey
NIH	National Institutes of Health
NORC	National Opinion Research Center
NPEC	National Postsecondary Education Cooperative
NPEFS	National Public Education Financial Survey
NPSAS	National Postsecondary Student Aid Study
NRC	National Research Coordinator (see TIMSS)
NSF	National Science Foundation
NSOPF	National Study of Postsecondary Faculty
NTID	National Technical Institute for the Deaf

OECD	Organization for Economic Cooperation and Development (see IEA Reading Literacy Study)
OERI	Office of Educational Research and Improvement (name changed to Institute of Education Sciences, or IES, in 2002)
OLDS	Oral Language Development Scale (see ECLS)
OMB	Office of Management and Budget
OPE	Office of Postsecondary Education (see IPEDS)

PAPI	Paper-and-pencil interviewing
PC CARP	computer statistical package
PEPS	Postsecondary Education Participants System (see IPEDS)
PEQIS	Postsecondary Education Quick Information System
PETS	Postsecondary Education Telephone System (see IPEDS); Postsecondary Education Transcript Study (see NLS-72 and HS&B)
PFI	Parent and Family Involvement in Education Survey (see NHES)
PIRLS	Progress in Reading Literacy Study (see IEA Reading Literacy Study)
PISA	Program for International Assessment (see IEA Reading Literacy Study)
PLS	Public Libraries Survey
PMSA	Primary Metropolitan Statistical Area
PPA	Program Participation Agreement (see IPEDS)
PPS	Probability Proportional to Size (a sampling method)
PROC IMPUTE	a regression-based imputation method developed for NCES by American Institutes for Research
PSS	Private School Universe Survey
PSU	Primary Sampling Unit

QED	Quality Education Data, Inc.

RDD	Random Digit Dialing (telephone survey technique)
ROTC	Reserve Officers Training Corps
RTD	Registered Time-to-Degree (see SED)
RTI	Research Triangle Institute

S	Fall Staff (see IPEDS)
SAQ	Self-Administered Questionnaire
SAS	computer statistical package; computer language designed specifically for the manipulation of statistical data
SASS	Schools and Staffing Survey
SAVD	School Associated Violent Death Study
SCS	School Crime Supplement
SD	Students with Disabilities
SDR	Survey of Doctorate Recipients (see SED)

SEA	State Education Agency (see CCD)
SED	Survey of Earned Doctorates
SES	Socioeconomic status
SFA	Student Financial Aid (see IPEDS)
SIF	School Information Form (see HST)
SIMS	Second International Mathematics Study (see TIMSS)
SIPP	Survey of Income and Program Participation (see NHES)
SISS	Second International Science Study (see TIMSS)
SLS	School Library Survey
SPSS	computer statistical package
SR	School Readiness Survey (see NHES)
SRIF	Student Record Information Form (see NLS-72)
SSD	School Safety and Discipline Survey (see NHES)
SSOCS	School Survey on Crime and Safety
StLA	State Library Agency Survey
SUDAAN	SUrvey DAta ANalysis, a computer statistical package
SUREG	a computer program used in HS&B
TCMA	Test-Curriculum Matching Analysis (see TIMSS)
TFS	Teacher Follow-up Survey
TIGER	Topologically Integrated Geographical Encoding and Referencing (see NALS)
TIMSS	Third International Mathematics and Science Study
TIMSS-R	Third International Mathematics and Science Study-Repeat
TRP	Technical Review Panel (see IPEDS)
TSA	Trial State Assessment (see NAEP)
TTD	Total Time-to-Degree (see SED)
USDA	United States Department of Agriculture
USU	Ultimate Sampling Unit
WesVar	computer statistical package
YALS	Young Adult Literacy Survey
YCI	Youth Civic Involvement Study (see NHES)
YRBS	Youth Risk Behavior Survey (see SCS)

Introduction

Since its inception, the National Center for Education Statistics (NCES) has been committed to the policy of explaining its statistical methods to its customers and of seeking to avoid misinterpretation of its published data. The reason for this policy is to assure customers that proper statistical standards and techniques have been observed, to guide them in the appropriate use of information from NCES, and to make them aware of the known limitations of NCES data.

This first edition of the *NCES Handbook of Survey Methods* continues this commitment by presenting current explanations of how each survey program in NCES obtains and prepares the data it publishes. NCES statistics are used for many purposes, and sometimes data well suited to one purpose may have limitations for another. This handbook aims to provide users of NCES data with the most current information necessary to evaluate the suitability of the statistics for their needs, with a focus on the methodologies for survey design, data collection, and data processing. It is intended to be used as a companion report to *Programs and Plans of the National Center for Education Statistics*, which provides a summary description of the type of data collected by each program at the Center.

LAYOUT OF HANDBOOK CHAPTERS

- Overview
- Uses of Data
- Key Concepts
- Survey Design
- Data Quality & Comparability
- Contact Information
- Methodology & Evaluation Reports

NCES Role and Organization

Among federal agencies collecting and issuing statistics, NCES is a general-purpose statistical collection agency in the broad field of education. The Center's data serve the needs of Congress, other federal agencies, national education associations, academic education researchers, business, and the general public. NCES is a component of the Institute of Education Sciences (IES), formerly the Office of Educational Research and Improvement (OERI), within the Department of Education.

Within NCES, the Statistical Standards Program, under the direction of the NCES Chief Statistician, provides expertise in statistical standards and methodology, technology, and customer service activities across subject matter lines. Specific survey programs of NCES have developed around subject matter areas. As a result, the rest of NCES is organized according to subject matter areas, with each of the survey programs falling under one of the following four NCES divisions:

- Assessment
- Early Childhood, International, and Crosscutting Studies
- Elementary/Secondary and Libraries Studies
- Postsecondary Studies

Organization of the Handbook

The handbook contains 28 chapters. Chapters 1 to 26 each focus on one of the 26 major NCES survey programs. To facilitate locating similar information for the various programs, the information in each of these chapters is presented in a uniform format with the following standard sections and headings:

1. *Overview.* This section includes a description of the purpose of the survey, the type of information collected in the survey, and the periodicity of the survey.
2. *Uses of Data.* This section summarizes the range of issues addressed by the data collected in the survey.
3. *Key Concepts.* This section provides the definitions of a few important concepts specific to the survey.
4. *Survey Design.* This section describes the target population, the sample design, the data collection and processing procedures, the estimation methods, and future plans for the survey. Note that the handbook does not include a list of the data elements collected by each survey. That information can be found in the survey questionnaires, electronic codebooks, or data analysis systems, many available through the NCES web site (http://nces.ed.gov). However, some general remarks about the data collected can be made here:
    - All race/ethnicity data are collected by Office of Management and Budget (OMB) standards.
    - All data on individuals can be disaggregated by sex.
    - All data on individuals can also be disaggregated by Black, White, and Other, and, for some surveys, data can be disaggregated by Hispanic and Asian/Pacific Islander.
    - All elementary/secondary student-level data collections include information on limited-English proficiency and student disability.
    - School-level data collections include information on programs and services offered.
5. *Data Quality and Comparability.* This section describes the appropriate method to use for estimating sampling error for sample surveys and also presents important findings related to nonsampling error such as coverage error, unit and item nonresponse error, and measurement error. In addition, this section provides summary descriptions of recent design and/or questionnaire changes as well as information on comparability of similar data collected in other studies.
6. *Contact Information.* This section lists the name of the main contact person for each survey along with a telephone number, e-mail address, and mailing address. Note that at NCES, telephone numbers are assigned according to survey program; staff members leaving one survey program for another have to change telephone numbers. To find out the current number for a particular staff member, see the NCES Staff Directory (http://nces.ed.gov/ncestaff/). To find out the current contacts for a particular survey program, please check the program's web site (NCES survey web site addresses are listed in appendix D).
7. *Methodology and Evaluation Reports.* This section lists the primary recent methodological reports for the survey. Use the NCES number provided to find a particular report through the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/). Each NCES survey Web site also contains a list of that survey's publications.

Note that some of the chapters include cautions to data users. The cautions usually appear in section 5, Data Quality and Comparability. For example, in chapter 5, section 5, caution is urged in the interpretation of change estimates between the 1991–92 and 1994–95 Teacher Follow-up Survey (TFS) because specific questions were not always worded the same in both TFS surveys. In chapter 11, section 5, users of Academic Library Survey data are reminded to be careful when comparing state estimates since nonresponse varies by state. These cautions are italicized throughout the report.

The first 26 chapters are organized under the following subject matter rubrics:

- Early Childhood Education Survey
    - Chapter 1: Early Childhood Longitudinal Study (ECLS)
- Elementary and Secondary Education Surveys
    - Chapter 2: Common Core of Data (CCD)
    - Chapter 3: Private School Universe Survey (PSS)
    - Chapter 4: Schools and Staffing Survey (SASS)
    - Chapter 5: SASS Teacher Follow-up Survey (TFS)
    - Chapter 6: National Education Longitudinal Study of 1988 (NELS:88)
    - Chapter 7: National Longitudinal Study of the High School Class of 1972 (NLS-72)
    - Chapter 8: High School and Beyond (HS&B) Longitudinal Study
- Library Surveys
    - Chapter 9: SASS School Library Survey (SLS)
    - Chapter 10: Public Libraries Survey (PLS)
    - Chapter 11: Academic Libraries Survey (ALS)
    - Chapter 12: State Library Agencies (StLA) Survey

BEST COPY AVAILABLE

  - Chapter 13: Federal Libraries and Information Centers Survey
- Postsecondary and Adult Education Surveys
  - Chapter 14: Integrated Postsecondary Education Data System (IPEDS)
  - Chapter 15: National Study of Postsecondary Faculty (NSOPF)
  - Chapter 16: National Postsecondary Student Aid Study (NPSAS)
  - Chapter 17: Beginning Postsecondary Students (BPS) Longitudinal Study
  - Chapter 18: Baccalaureate and Beyond (B&B) Longitudinal Study
  - Chapter 19: Survey of Earned Doctorates (SED)
- Educational Assessment Surveys
  - Chapter 20: National Assessment of Educational Progress (NAEP)
  - Chapter 21: Third International Mathematics and Science Study (TIMSS)
  - Chapter 22: IEA Reading Literacy Study (IEA)
  - Chapter 23: National Adult Literacy Survey (NALS)
  - Chapter 24: International Adult Literacy Survey (IALS)
- Household Surveys
  - Chapter 25: National Household Education Surveys (NHES) Program
  - Chapter 26: Current Population Survey—October and September Supplements (CPS)

Chapters 27 and 28 cover multiple surveys or survey systems. The format is similar to that for chapters 1 to 26, but is somewhat abbreviated to allow adequate coverage of multiple surveys within each chapter.

- Small Special-Purpose NCES Surveys
  - Chapter 27: Fast Response Surveys
    - Fast Response Survey System (FRSS)
    - Postsecondary Education Quick Information System (PEQIS)
  - Chapter 28: Other NCES Surveys and Studies
    - School Crime Supplement (SCS)
    - School Survey on Crime and Safety (SSOCS)
    - High School Transcript (HST) Studies
    - Library Cooperatives Survey (LCS)
    - Civic Education Study (CivEd)

To avoid repetition within the handbook, some of the statistical terms and procedures that are referred to in multiple chapters of the handbook are defined in **Appendix A, Glossary of Statistical Terms**.

Appendix B describes the various ways in which NCES publications and data files may be obtained. It also provides the reader with information on how to obtain a license for restricted-use data files.

Appendix C provides a list of the web-based and standalone tools for use with each of the NCES surveys.

Appendix D contains a list of the web site addresses for each of the NCES surveys.

Appendix E contains an index.

Chapter 1: Early Childhood Longitudinal Study (ECLS)

1. OVERVIEW

The Early Childhood Longitudinal Study program is one of four active longitudinal surveys sponsored by NCES, and the first to provide data about young children. The ECLS program has been designed to include two overlapping cohorts: a birth cohort and a kindergarten cohort. The birth cohort component (ECLS-B) will follow a sample of children born in 2001 from birth through the 1st grade, while the kindergarten component (ECLS-K) will follow a sample of children who were in kindergarten in the 1998–99 school year through the 5th grade. ECLS will provide a comprehensive and reliable data set with information about the ways in which children are prepared for school and how schools and early childhood programs affect the lives of the children who attend them.

EARLY CHILDHOOD LONGITUDINAL SAMPLE SURVEY: BIRTH COHORT AND KINDERGARTEN COHORT

ECLS collects data from:

- Children
- Parents/guardians
- Child care providers and preschool teachers
- Teachers
- School administrators

Purpose

ECLS provides national data on (1) children's status at birth and at various points thereafter; (2) children's transitions to nonparental care, early education programs, and school; and (3) children's experiences and growth through the 5th grade. These data enable researchers to test hypotheses about the effects of a wide range of family, school, community, and individual variables on children's development, early learning, and early performance in school.

Components

ECLS has two cohorts—the *kindergarten* cohort study (ECLS-K) and the *birth* cohort study (ECLS-B)—and each of these has its own components.

Kindergarten cohort study. ECLS-K collects data from children, parents, classroom teachers, special education teachers, school administrators, and student records. The various components are described below.

Direct child assessments. The direct child assessments consist of three cognitive domains (reading, mathematics, and general knowledge); a psychomotor assessment (fall kindergarten only), including fine and gross motor skills; and height and weight measurements. An English language proficiency screener, the Oral Language Development Scale (OLDS), is administered if the school records indicate that the child's home language is not English. The child has to demonstrate a certain level of English proficiency to be administered the ECLS-K cognitive assessment in English. If a child speaks Spanish at home and does not have the English skills required by the ECLS-K battery, the child is administered a Spanish version of the OLDS, and the mathematics and psychomotor assessments are administered in Spanish. Each cognitive assessment domain subtest includes a routing test (to determine a child's approximate skills) and level tests.

Parent interviews. Parents/guardians are asked to provide key information about their children on subjects such as family demographics (e.g., age, relation to child, race/ethnicity), family structure (household members and composition), parent/guardian involvement, home educational activities, childcare experience, child health, parental/guardian education and employment status, and their child's social skills and behaviors.

Classroom teacher questionnaire. In the base year, all kindergarten teachers in the ECLS-K schools were asked to provide information on their educational backgrounds, teaching practices, experiences, and the classroom settings where they taught. Kindergarten teachers who taught ECLS-K-sampled children also completed a child-specific questionnaire that collected information on each child's social skills and approaches to learning, academic skills, and education placements. In the 1st grade and later waves of the study, only teachers of the sampled children are included.

Special Education Teacher Questionnaire. The special education teacher questionnaires were introduced in the spring data collection. ECLS-K supervisors reviewed accommodation and inclusion information for children who received special education services. During the preassessment visit, field supervisors specified primary special education teachers of sampled children and listed special education staff working with each child (e.g., speech pathologists, reading instructors, audiologists). These questionnaires were given to special education teachers who taught sampled children. If a child received special education services from more than one special education teacher, a field supervisor determined the child's primary special education teacher. Items in the special education teacher questionnaires addressed topics such as the child's disability, Individual Education Program goals, the amount and type of services used by sampled students, and communication with parents and general education teachers.

School Administrator Questionnaire. School administrators are asked about school characteristics (e.g., school type, enrollment, and student body composition), school facilities and resources, community characteristics and school safety, school policies and practices, school-family-community connections, school programs for special populations, staffing and teacher characteristics, school governance and climate, and their own characteristics.

Student Records Abstract. School staff members are asked to complete a student records abstract form for each sampled child after the school year closed. These instruments were used to obtain information about the child's attendance record, the presence of an individualized education plan, the type of language or English proficiency screening that the school used, and (in the kindergarten year collection) whether the child participated in Head Start prior to kindergarten. A copy of each child's report card was also requested.

School Facilities Checklist. The checklist collects information about the (1) availability and condition of the selected schools' facilities such as classrooms, gymnasiums, toilets, etc.; (2) presence and adequacy of security measures; (3) presence of environmental factors that may affect the learning environment; and (4) overall learning climate of the school. An additional set of questions on portable classrooms was added to the spring-1st-grade data collection.

Birth cohort study. The ECLS-B, implemented in October 2001, is designed to study children's early learning and development from birth through 1st grade. Over the course of the study, data will be collected from multiple sources, including birth certificates, children, parents, nonparental care providers, teachers, and school administrators. These components are described below.

Birth certificates. These records provide information on the date of birth, child's sex, parents' education, parents' race and ethnicity (including Hispanic origin), mother's marital status, mother's pregnancy history, prenatal care, medical and other risk factors during this pregnancy and complications during labor and birth, and child's health characteristics (such as congenital anomalies and abnormal conditions of the baby and the baby's APGAR score).

Parent/guardian interviews. A parent/guardian interview is conducted in the children's home at each data collection point to capture information about the children's early health and development, their experiences with family members and others, the parents/guardians as caregivers, the home environment, and the neighborhood in which they live. In most cases, the parent/guardian interviewed is the child's mother or female guardian.

Child assessments. Beginning at 9 months, children participate in activities designed to measure important developmental skills in the cognitive, social, emotional, and physical domains. ECLS-B uses adapted forms of the Bayley Scales for Infant Development (BSID-II) and the Nursing Child Assessment Teaching Scale (NCATS). The children's height, weight, and middle upper arm

BEST COPY AVAILABLE

circumference are assessed at the 9-month home visit. In addition, during the home visit children's psychomotor skills and emotion regulation will be assessed. At the 18-month home visit, the Massey Attachment Sort (MAS) will be used to assess children's levels of attachment with their caregivers. (For further details, see Assessment Design.)

Care provider and preschool teacher interviews. Individuals and organizations who provide regular care for a child will be interviewed with the permission of the child's parents. They will be asked about their backgrounds, teaching practices, and experience, the children in their care, and children's learning environments. This information will be collected when the children are 18 months of age and again at 48 months.

School administrator/teacher questionnaires. Once the children enter formal schooling, school administrators and teachers will provide information on the physical and organizational characteristics of their schools and on the schools' learning environments, educational philosophies, and programs. Teachers will also provide information on the classroom, and they represent important potential sources of information about children's cognitive and social development.

Father questionnaire. Fathers will complete a self-administered questionnaire reviewing the particular role fathers play in the development of their children, providing information about children's well-being and the activities fathers engage in with their children as well as key information about themselves as caregivers. This information will be collected when the children are 9 and 18 months old and at least two additional times during the study.

Periodicity
Each of the ECLS cohorts has its own follow-up schedule.

The ECLS-K schedule is for data collection in the fall and spring of the kindergarten year (1998–99), a 30 percent fall 1st-grade subsample (1999), and a full sample for spring of the 1st (2000), 3rd (2002), and 5th (2004) grades.

The ECLS-B schedule is for data collection at 9 months (2001–02), 18 months (2002–03), 30 months (2003–04), 48 months (2005), kindergarten (2006 and 2007), and 1st grade (2007 and 2008). Note that because of age requirements for school entry, children sampled in ECLS-B will be entering kindergarten, and thus 1st grade, in two different years.

2. USES OF DATA

ECLS-K provides information critical to establishing policies that can respond sensitively and creatively to diverse learning environments. In addition, ECLS-K will enable researchers to study how a wide range of family, school, community, and individual variables affect early childhood success in school. The information collected during the kindergarten year serves as baseline data to examine how schooling shapes later development. The longitudinal nature of the study will enable researchers to study children's reading achievement, growth in mathematics, and knowledge of the physical and social worlds in which they live. It will also permit researchers to relate trajectories of growth and change to variations in children's school experiences in kindergarten and the early grades.

Like the kindergarten cohort study, ECLS-B has two goals, descriptive and analytic. The study will provide descriptive data on children's health status at birth; children's experiences in the home, nonparental care, and school; and children's development and growth through 1st grade. The study will also collect data that can be used to explore the relationships between children's developmental outcomes and their family, health care, nonparental care, school, and community. Data collected during the first year of life (around 9 months) will serve as a baseline for examining how children's home environment, health status, health care, and early childcare and education shape their development. The longitudinal nature of the study will enable researchers to study children's physical, social, and emotional growth and to relate trajectories of growth and change to variations in children's experience.

3. KEY CONCEPTS

IRT scale scores. These scores are overall, criterion-reference measures of status at a point in time. They are useful in identifying cross-sectional differences among subgroups in overall achievement level and provide a summary measure of achievement useful for correlations analysis with status variables. The IRT scale scores are used as longitudinal measures of overall growth. Gain scores may be obtained by subtracting children's scale scores at two points in time.

Standardized scores (T-scores). These scores provide norm-referenced measurements of achievement; that is, estimates of achievement level relative to the population as a whole. A high mean T-score for a particular subgroup indicates that the group's performance is high in comparison to other groups. A change in mean T-scores over time reflects a change in the group's status with respect to other groups. In other words, they provide information on status compared to children's peers.

Proficiency probability scores. These scores are criterion-referenced measures of proficiency in specific skills. Because each proficiency score targets a particular set of skills, they are ideal for studying the details of achievement. They are useful as longitudinal measures of change because they show not only the extent of gains but also where on the achievement scale the gains are taking place. The following proficiencies were identified in the reading and mathematics assessments:

Reading proficiencies:

- Letter recognition: identifying upper- and lower-case letters by name
- Beginning sounds: associating letters with sounds at the beginning of words
- Ending sounds: associating letters with sounds at the end of words
- Sight words: recognizing common words by sight
- Comprehension of words in context: reading words in context

Mathematics proficiencies:

- Number and shape: identifying some one-digit numerals, recognizing geometric shapes, and one-to-one counting of up to 10 objects
- Relative size: reading all single-digit numerals, counting beyond 10, recognizing a sequence of patterns, and using nonstandard units of length to compare objects
- Ordinality, sequence: reading two-digit numerals, recognizing the next number in a sequence, identifying the ordinal position of an object, and solving a simple word problem
- Addition/subtraction: solving simple addition and subtraction problems
- Multiplication/division: solving simple multiplication and division problems and recognizing more complex number patterns

Race/ethnicity. New Office of Management and Budget guidelines were followed under which a respondent could select one or more of five dichotomous race categories. In addition, a sixth dichotomous variable was created for those who simply indicated that they were multiracial without specifying the race. Each respondent additionally had to identify whether the child was Hispanic. Using the six dichotomous race variables and the Hispanic ethnicity variable, a race/ethnicity composite variable was created. The categories were: White, non-Hispanic; Black or African-American, non-Hispanic; Hispanic, race specified; Hispanic, no race specified; Asian; Native Hawaiian or other Pacific Islander; American Indian or Alaskan Native; and more than one race specified, non-Hispanic.

Socioeconomic scale. The socioeconomic scale (SES) variable was computed at the household level for the set of parents who completed the parent interview in ECLS-K. The SES variable reflects the socioeconomic status of the household at the time of data collection. The components used to create the SES variable were: father/male guardian's education, mother/female guardian's education, father/male guardian's occupation, mother/female guardian's occupation, and household income. Each parent's occupation was scored using the average of the 1989 General Social Survey prestige scores for the 1980 Census occupational category codes that correspond to the ECLS-K occupation code.

4. SURVEY DESIGN

Target Population

Representative samples of kindergartners and babies will be studied longitudinally for 6 or more years. Kindergarten children enrolled during the 1998–99 school year will be the baseline for the ECLS-K cohort, babies born during 2001 will consist of the baseline for the ECLS-B cohort.

Sample Design

The sampling design is discussed separately for the kindergarten and birth cohorts.

Kindergarten Cohort (ECLS-K). ECLS-K is following a nationally representative cohort of children from kindergarten through 5th grade.

Base Year Survey. A nationally representative sample of 22,782 children enrolled in 1,277 kindergarten programs during the 1998–99 school year was sampled for participation in the study. These children were selected from both public and private kindergartens, offering both full-day and part-day programs. The sample was designed to

BEST COPY AVAILABLE

support separate estimates of public and private school kindergartners; Black, Hispanic, White, and Asian and Pacific Islander (API) children; and children grouped according to socioeconomic status.

The sample design for ECLS-K was a dual-frame, multistage sample. First, 100 primary sampling units (PSUs) were selected from an initial frame of 1,404 PSUs, representing counties or groups of contiguous counties. The 24 PSUs with the largest measures of size (where the measure of size is the number of 5-year-olds, taking into account a factor for oversampling 5-year-old APIs) were designated as certainty selections and were set aside. The remaining PSUs were partitioned into 38 strata of roughly equal measure of size. The frame of noncertainty PSUs was first sorted into eight superstrata by metropolitan statistical area (MSA) status and by Census region. Within the four MSA superstrata, the variables used for further stratification were race/ethnicity (high concentration of API, Black, or Hispanic), size of class, and 1988 per capita income. Within the four non-MSA superstrata, the stratification variables were race/ethnicity and per capita income. Two PSUs were selected from each noncertainty stratum using Durbin's Method. This method selects two first-stage units per stratum without replacement, with probability proportional to size and a known probability of inclusion. The Durbin method was used because it allows variances to be estimated as if the units were selected with replacement.

The school selection occurred within these PSUs. Public schools were sampled from a public school frame (the 1995–96 Common Core of Data—CCD), and private schools were sampled from a private school frame (the 1995–96 Private School Survey—PSS). The school frame was freshened in the spring of 1998 to include newly-opened schools that were not included in the CCD and PSS and schools that were in the CCD and PSS but did not offer kindergarten according to these sources. A school sample supplement was selected from the freshened frame. In fall 1998, approximately 23 kindergarten children were selected on average from each of the sampled schools. API children and private schools were oversampled.

Fall-1st grade. This study was a design enhancement whose goal was to enable researchers to measure the extent of summer learning loss and the factors that contribute to such loss and to better disentangle school and home effects on children's learning. Data collection was limited to 26.7 percent of the base year children in 30 percent of the ECLS-K originally sampled schools; that is, a total of 5,650 (4,446 public and 1,204 private) children and 311 schools (228 public and 83 private). Data collection was attempted for every eligible child (i.e., a base year respondent) found still attending the school in which he or she had been sampled during kindergarten. To contain the cost of collecting data for a child who transferred from the school in which he or she was originally sampled, a random 50 percent of children were flagged to be followed for fall-1st-grade data collection in the event that they had transferred.

Spring-1st grade. This data collection targeted all base year respondents. In addition, the spring student sample was freshened to include current 1st graders who had not been enrolled in kindergarten in 1998–99 and, therefore, had no chance of being included in the ECLS-K base year kindergarten sample. While all students still enrolled in their base year schools were recontacted, only a 50 percent subsample of base year sampled students who had transferred from their kindergarten school was followed for data collection. The sample of base year respondents numbered 18,084 (14,248 public and 3,836 private) children. Student freshening brought 165 1st graders into the ECLS-K sample.

Birth Cohort (ECLS-B). ECLS-B sampled approximately 16,000 babies born in the year 2001. The sample includes children from different racial/ethnic and socioeconomic backgrounds. Chinese children, other API children, moderately low birth weight children (1500–2500 grams), very low birth weight children (under 1500 grams), and twins were oversampled. There was also a special supplemental component to oversample American Indian children (with an initial sample size of 1,299).

The ECLS-B sample design consists of a two-stage sample of PSUs and children born in the year 2001 within sampled PSUs. The PSUs are MSAs, counties, or groups of counties, and were selected with probability proportional to a function of the expected number of births occurring within the PSU in 2001. Births were sampled by place of occurrence, rather than by place of current residence. As a result, a different PSU sample had to be selected from the PSU sample used in ECLS-K, which uses residence-based population data. Within the sampled PSUs, children born in the year 2001 were selected by systematic sampling from birth certificates using the National Center for Health Statistics (NCHS) vital statistics record system. The sample was selected on a flow basis, beginning with January 2001 births (who were first assessed 9 months later, in October 2001). Approximately equal numbers of infants were sampled from each month.

Different sampling rates were used for births in different subgroups, as defined by race/ethnicity, birth weight, and plurality (that is, whether or not the sampled newborn is a twin).

The sample of American Indian newborns drew from additional PSUs in three states that are not included in the ECLS-B main study. Because these three additional states would not allow use of their birth certificate information, an alternate frame was used. A hospital sample frame was chosen based on an evaluation of available sample frames.

Due to state-imposed operational restrictions and passive and active consent procedures, certain PSUs had low expected response rates. For states where expected response rates were only slightly lower than planned, a larger sample was selected in order to achieve adequate numbers of respondents. Substitutions were made for PSUs in states where very low response rates were expected. The original PSU was matched with potential substitute PSUs on the criteria of median income, percent of newborns in poverty, percent of minority newborns, population density, and birth rate. American Indian PSUs were also matched on tribal similarity. A Mahalonobis distance measure of similarity was used to create initial rankings. Sampling rates from the original PSU were applied within the substitute PSU to obtain the original expected yield.

Assessment Design

The design of the ECLS assessments is discussed separately for the kindergarten and birth cohorts.

Kindergarten Cohort (ECLS-K). The design of the ECLS-K assessment was guided by the domain assessment framework proposed by the National Education Goals Panel's Resource Group on School Readiness. A critical component of ECLS-K is the assessment of children along a number of dimensions, such as physical development, social and emotional development, and cognitive development. These domains were chosen because of their importance to success in school. ECLS-K will monitor the status and growth of its children along these domains:

- *Physical and psychomotor development:* Children's height and weight will be measured at each data collection period in ECLS-K. In the fall of kindergarten, kindergartners were asked to demonstrate their fine and gross motor skills through activities such as building a structure using blocks, copying shapes, drawing figures, balancing, hopping, skipping, and walking backwards. Parents and teachers report on other related issues such as general health, nutrition, and physical activity.
- *Social and emotional development:* ECLS-K assessments of social and emotional development focus on the skills and behaviors that contribute to social competence. Aspects of social competence include social skills (e.g., cooperation, assertion, responsibility, self-control) and problem behaviors (e.g., impulsive reactions, verbal and physical aggression). Parents and teachers are the primary sources of information on children's social competence and skills, at least from kindergarten through 2nd grade. The measurement of children's social and emotional development at grades 3 and 5 will include instruments completed by the children themselves along with data reported by parents and teachers.
- *Cognitive development:* ECLS-K focuses on three broad areas of competence: language and literacy, mathematics, and general knowledge of the social and physical worlds. The skills measured in each of these domains are a sample of the typical and important skills that are taught in American elementary schools and that children are expected to learn in school. ECLS-K was developed to describe the behaviors, skills, and knowledge within broad cognitive domains that are most relevant to school curricula at each grade level and to measure children's growth from kindergarten to 5th grade. The ECLS-K assessment framework was based on current curricular domain frameworks for reading, mathematics, science, and social studies, as well as assessment frameworks such as the National Assessment of Educational Progress. (See chapter 20.)

These assessments were developed after extensive field testing and analysis. The final items were selected based on their psychometric properties and content relevance. The measure of language and literacy competency includes vocabulary comprehension, listening and reading comprehension, and basic skills (e.g., knowledge of the alphabet, phonetics, print recognition and orientation, and sight vocabulary). The mathematics subdomain measures the knowledge and skills necessary to solve mathematical problems and reason with numbers. The items measuring children's quantitative and analytic skills in kindergarten and 1st grade include recognizing numbers, counting, comparing and ordering numbers, and solving word problems. Other measures of mathematical concepts include recognizing and solving problems involving graphic and numeric patterns and geometric relationships. Items involving the interpretation of picture graphs measure beginning analysis and statistics skills. Children's knowledge and skills in the natural and social sciences are measured in the general knowledge subdomain. The contents of this subtest, classified as science and social sciences, survey children's knowledge and understanding of relevant concepts.

Each direct child domain subtest consists of a routing test and level tests. All children are first administered a short routing test of domain-specific items having a broad range of complexity or difficulty levels. Performance on the routing test is used to determine the appropriate level assessment form that will be administered next to the child. The use of multilevel forms for each domain subtest minimizes the chances of administering items that are all very easy or all very difficult for the child. Children demonstrate their competency in these domains through one-on-one, untimed sessions with a trained child assessor. If necessary, the session can take place over multiple periods.

Birth Cohort (ECLS-B). The ECLS-B direct child assessment relies on instruments considered "gold standards" in the field. However, adaptations have been necessary to take these instruments from a laboratory or clinic setting to a home setting. The ECLS-B child assessment, therefore, is designed for ease of and flexibility in administration while at the same time being psychometrically and substantively sound. The key instruments are a shortened research edition of the BSID-II, NCATS, and an attachment measure—MAS.

- *Cognitive development and fine and gross motor skills:* BSID-II is considered the gold standard for assessing early childhood development (ages 1 to 42 months). Children's cognitive development, as well as their receptive and expressive language skills, are assessed through the mental scale of the BSID-II. Children retrieve hidden toys and look at pictures books, and their production of vowel-consonant combinations is noted. Fine and gross motor skills are assessed through the motor scale of the BSID-II. Children grasp small objects and are observed crawling and walking. The Bayley assessment was originally expected to take about 20 minutes. However, a field test of the 9-month ECLS-B data collection revealed that it actually required an average of 40 minutes to complete. As a result, modifications were implemented to the original BSID-II. The ECLS-B contractor, Westat, worked with experts to identify a reduced-item set that can be administered in less time and can produce reliable, valid scores equivalent to the full set of Bayley items. The ECLS-B reduced-item Bayley for 9-month-olds takes approximately 25 minutes to administer.

- *Parent-child interaction:* NCATS is designed to assess parent-child interaction (ages 0 to 36 months). Parents are asked to teach their child a task that she or he cannot do from a standard list using NCATS materials. Tasks include turning pages of a book and stacking blocks. The interaction is videotaped and later coded along six subscales. The teaching scale provides information on child cues, parent responsiveness, and the fostering of socio-emotional and cognitive growth. It captures variables that are precursors to later social and cognitive development, such as attachment and language.

- *Attachment with caregivers:* The Strange Situation and the Attachment Q-Sort (AQS) are the commonly used measures for assessing and discussing toddlers' attachment relationships. These measures require a significant amount of time to complete and are fairly complex for a field staff. MAS is an alternative to the laboratory-based Strange Situation measure, developed exclusively for ECLS-B. It uses the Method of Successive Sorts (MOSS), which is considered to be operationally easier than the Q-sort. MAS features 39 AQS items, which have been edited to an elementary reading level. Parents and field staff work with a deck of cards and sort descriptions of parent/child behavior for how much it is like the child. Card descriptions include scenarios to assess the child's proximity to the parent and exploration behavior and the occurrence of differential responsiveness. Aspects of children's affect, sociability, and independence are also assessed. MAS can be completed by respondents and field staff from different backgrounds, and it can be completed in less than 10 minutes.

Data Collection and Processing

ECLS-K compiles data from four primary sources: children, children's parents/guardians, teachers, and school administrators. Data collection began in fall 1998 and will continue through spring 2004. Westat has collected the kindergarten and 1st-grade data. ECLS-B compiles data from multiple sources, including administrative records, children, parents, nonparental care providers, teachers, and school administrators. Data collection began in 2001 and will continue through 2008. Self-administered questionnaires, one-on-one assessments, and telephone or personal interviews will be used to collect the data. Westat is the 9- and 18-month data collection contractor.

Reference dates. For ECLS-K, baseline data for the fall were obtained from September to December 1998. For ECLS-B, baseline data was collected from October 2001 through December 2002.

Data collection. ECLS-K and ECLS-B are discussed separately.

Kindergarten Cohort (ECLS-K). The data collection schedule for ECLS-K is based on a desire to capture information about children as critical events and transitions are occurring rather than measuring these events retrospectively. A large-scale field test of the kindergarten and 1st-grade assessment instruments and questionnaires was conducted in 1995–96. This field test was used primarily to collect psychometric data on the ECLS-K assessment item pool and to evaluate questions in the different survey instruments. Data from this field test were used to develop the first- and second-stage tests for the ECLS-K kindergarten and 1st-grade direct cognitive assessment battery and to finalize the parent, teacher, and school administrator instruments. A pilot test of the systems and procedures, including field supervisor and assessor training, was conducted in April and May 1998 with 12 elementary schools in the Washington, DC metropolitan area. Modifications to the data collection procedures, training programs, and systems were made to improve efficiency and reduce respondent burden. Modifications to the parent interview to address some issues raised by pilot test respondents were also made at this time.

Data on the kindergarten cohort were collected twice during the base year of the study—once in the beginning (fall) and once near the end (spring) of the 1998–99 school year. The fall 1998 data collection obtained baseline data on children prior to their exposure to the influences of school, providing measures of the characteristics and attributes of children as they entered formal school for the first time. The data collected in spring 1999, together with the data from the beginning of the school year, are used to examine children's first encounter with school. Data were collected from the child, the child's parents/guardians, and teachers. For the fall 1998 and spring 1999 collections, all child assessment measures were obtained through untimed CAPI, administered one-on-one from the assessor to child. Most of the parent data were collected through CATI, though some of the interviews were collected through CAPI when respondents did not have a telephone or were reluctant to be interviewed by telephone. All kindergarten teachers with sampled children were asked to fill out two self-administered questionnaires providing information on themselves and their teaching practices. For each of the sampled children they taught, the teachers also completed a child-specific questionnaire. In addition, school staff members were asked to complete a student record abstract after the school year closed; they were reimbursed five dollars for every student record abstract they completed.

In fall 1999—when most of the kindergarten cohort had moved on to 1st grade—data were collected from a 30 percent subsample of the cohort. School administrators were contacted in late summer 1999, and parental consents were reviewed (and re-obtained, if necessary). The direct child assessment was administered during a 12-week field period (September–November 1999). It was normally conducted in a school classroom or library and took approximately 50 to 70 minutes per child. As in the spring-kindergarten data collection, children with a language other than English in the home who did not take the English ECLS-K battery in the prior were first administered the OLDS to determine what path was followed in fall-1st grade. Children who fell below the cut score for the OLDS and whose language was Spanish were administered a Spanish language version of the OLDS and the ECLS-K mathematics assessment, and had their height and weight measured. Children who fell below the cut score and whose language was other than Spanish had only their height and weight measured. The parent interview was administered between early September and mid-November 1999; it averaged 35 minutes, and was conducted primarily by telephone.

Spring data collection included direct child assessments, parent interviews, teacher and school questionnaires, student records abstracts, and the facilities checklist. As in other rounds, the child assessments were administered with CAPI assistance (March–June 2000), while both CATI and CAPI were used for the parent interview (March–July 2000). Self-administered questionnaires were used to gather information from teachers, school administrators, and student records (March–June 2000, but field staff prompted by telephone for the return of these materials through October 2000). Teachers were reimbursed seven dollars for each child rating they completed, and school staff were reimbursed seven dollars for every student record abstract they completed.

A continuous quality assurance process has been applied to all data collection activities. Data collection quality control efforts begin with the development and testing of the CATI and CAPI applications and the FMS. As these applications are programmed, extensive testing of the system is conducted. Quality control processes continue with the development of field procedures that maximize cooperation and thereby reduce the potential for nonresponse bias. Quality control activities are also practiced during training and data collection. During the original assessor training, field staff practiced conducting the parent interview in pairs and practiced the direct child assessment with kindergarten children brought to the training site for this purpose. In later data collection periods, experienced staff use a home study training package while new staff are trained in classroom sessions.

BEST COPY AVAILABLE

When the fieldwork begins, field supervisors observes each assessor conducting child assessments and makes telephone calls to parents to validate the interview. Field managers also make telephone calls to the schools to collect information on the school activities for validation purposes.

Birth Cohort (ECLS-B). A field test of ECLS-B instruments and procedures was conducted in the fall of 1999. The design featured many different tasks. For example, while in the home, a field staff member had to complete approximately eleven discrete tasks, and each task had special skill requirements. Early in the field test, NCES and the ECLS-B contractor found several problems regarding the complexity of the home visit: while separately no one task was difficult, the total data collection protocol was complex, so it was necessary to simplify these tasks in order to reduce the burden on field staff and to ensure the reliable and valid administration of all tasks. As a result, several modifications were made to the original data collection design.

A second field test of ECLS-B instruments and procedures began in September 2000. A new sample was drawn consisting of 1,062 children born between January and April 2000. Home visits were conducted when the children were 9 months old and again when the children were 18 months of age. Results from this field test indicated that the changes to the design that resulted from the first field test were successful.

The ECLS-B schedule calls for information to be gathered on the babies and from the parents during an in-home visit. The children's mothers or primary providers participate in the 9-month and 18-month interviews. Fathers answer a set of questions regarding their involvement in their children's lives when the babies are 9 months of age. At the 18-month data collection point, additional information is collected in a telephone interview with the childcare provider (when applicable), and fathers are again asked to answer questions about their involvement with their children. ECLS-B uses adapted forms of BSID-II, NCATS, and MAS.

ECLS-B uses NCATS at the 9- and 18-month data collections. ECLS-B is videotaping NCATS, although it is more typical for a health or social service professional to complete NCATS via live coding (i.e., while the interaction occurs). While the interaction lasts only about 5 minutes, the ECLS-B field staff needs to observe and score 73 items of parent and child behavior. Given the other tasks the field staff must learn and complete, live coding would limit the number of scales that could realistically be used, thereby reducing the amount of information that can be gathered. The videotapes will be coded along all scales.

In addition to the parent/guardian and childcare-provider interviews, school administrators and teachers will provide information on the physical and organizational characteristics of their schools and on the schools' learning environments, educational philosophies, and programs. Teachers also represent important potential sources of information about children's development.

Editing. Within the CATI/CAPI instruments, ECLS-K respondent answers were subjected to both "hard" and "soft" range edits during the interviewing process. Responses outside the soft range of reasonably expected values were confirmed with the respondent and entered a second time. For hard-range items, out-of-range values were usually not accepted. If the respondent insisted that a response outside the hard range was correct, the assessor could enter the information in a comments data file. Data preparation and project staff reviewed these comments. Out-of-range values were accepted if the comments supported the response.

Consistency checks were also built into the CATI/CAPI data collection. When a logical error occurred during a session, the assessor saw a message requesting verification of the last response and a resolution of the discrepancy. In some instances, if the verified response still resulted in a logical error, the assessor recorded the problem either in a comment or on a problem report.

The overall data editing process consisted of running range edits for soft and hard ranges, running consistency edits, and reviewing frequencies of the results.

Estimation Methods

Data are weighted to compensate for differential probabilities of selection at each sampling stage and to adjust for the effects of nonresponse. A hot-deck imputation methodology is used to impute for missing values of all components of the SES in the ECLS-K study.

Weighting. Several sets of weights were computed for each of the four rounds of data collection (fall-kindergarten, spring-kindergarten, fall-1st grade, and spring-1st grade). Longitudinal weights were also computed for children with data from multiple rounds of the study. Unlike surveys that have only one type of survey instrument aimed at one type of sampling unit, the ECLS-K is a complex study with multiple types of sampling units, each having its own survey instrument. Each type of unit was selected into the sample through a different mechanism: children were sampled directly through a sample of schools; parents of the sampled children were automatically included in the survey; all kindergarten teachers in the sampled schools were included; special education teachers were in the sample if they taught any of the sampled children. Each sampled unit had its own survey instrument: children were assessed directly using a series of cognitive and physical assessments; parents were interviewed with a parent instrument; teachers filled out at least two different types of questionnaires depending on the round of data collection and on whether they were regular or special education teachers; school principals reported their school characteristics using the school administrator questionnaire. The stages of sampling in conjunction with the different nonresponse level at each stage and the diversity of survey instruments require that multiple sampling weights be computed for use in analyzing the ECLS-K data.

Essentially, weights are driven by three factors: (1) how many points in time would be used in analysis (e.g., longitudinal or cross-sectional); (2) what level of analysis would be conducted (e.g., child, teacher, or school); and (3) what source of data is used (e.g., child assessment, teacher questionnaire, parent questionnaire).

In general, weights were computed in two stages. In the first stage, base weights were computed. They are the inverse of the probability of selecting the unit. In the second stage, base weights were adjusted for nonresponse. Nonresponse adjustment cells were generated using variables with known values for both respondents and nonrespondents. Analyses using the Chi-squared Automatic Interaction Detector (CHAID) were conducted to identify variables most highly related to nonresponse. Once the nonresponse cells were determined, the nonresponse adjustment factors are the reciprocals of the response rates within the selected nonresponse cells.

The base weight for each school is the inverse of the probability of selecting the PSU multiplied by the inverse of the probability of selecting the school within the PSU. The base weights for eligible schools are adjusted for nonresponse, made separately for public and private schools.

The base weight for each child in the sample is the school nonresponse-adjusted weight for the school attended, multiplied by a poststratified within-school student weight (total number of students in the school divided by the number of students sampled in the school). The poststratified within-school weight was calculated separately for API and non-API children because different sampling rates were used for these two groups. Within a school, all API children have the same base weights and all non-API children have the same base weights. The parent weight, which is the weight used to produce ECLS-K estimates, is the base child weight adjusted for nonresponse to the parent interview. Again, these adjustments were made separately for public and private schools.

Scaling. Item Response Theory (IRT) was employed to calculate scores that could be compared regardless of which second stage form a student took. The items in the routing test, plus a core set of items shared among the different second stage forms, made it possible to establish a common scale.

Imputation. SES component variables were computed in the base year and spring-1st grade ECLS-K. The percentages of missing data for the education and occupation variables were small (2 to 11 percent in the base year, 4 to 8 percent in spring-1st grade); however, the household income variable had higher missing rates (28.2 percent missing data in the base year and 11 to 33 percent in spring-1st grade, depending on whether a detailed income range or the exact household income was requested). A standard (random selection within class) hot-deck imputation methodology was used to impute for missing values of all SES components in both years, although the procedure used in spring-1st grade differed in that the initial step in the imputation procedure was to fill in missing values from information gathered during an earlier interview with that parent, if one had taken place.

The SES component variables were highly correlated so a multivariate analysis was more appropriate for examining the relationship of the characteristics of donors and nonrespondents. CHAID was used to divide the data into cells based on the distribution of the variable to be imputed, in addition to analyzing the data and determining the best predictors.

The variables were imputed in sequential order and separately by type of household. For households with both parents present, the mother's and father's variables were imputed separately. If this was not the case, an "unknown" or missing category was created as an addition level for the CHAID analysis. As a rule, no imputed value was used as a donor. In addition, the same donor was not used more than two times. The order of the imputation for all the variables was from the lowest percent missing to the highest. Occupation imputation involved two steps.

First, the labor force status of the parent was imputed, whether the parent was employed or not. Then the parent's occupation was imputed only for those parents whose status was identified as employed either through the parent interview or the first imputation step. The variable for income was imputed last using a three-stage procedure, where if a respondent provided partial information about income, this was used in the imputation process.

Future Plans

The ECLS-B cohort may be followed beyond 1st grade. Whether this is feasible or affordable will be evaluated over the life of the study.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

The estimators of sampling variances for ECLS statistics take the ECLS complex sample design into account. Both replication and Taylor Series methods have been developed. The paired jackknife replication method using 90 replicate weights can be used to compute approximately unbiased estimates of the standard errors of the estimates. (The fall 1st-grade subsample uses 40 replicates.) When using the Taylor Series method, a different set of stratum and first-stage unit identifiers should be used for each set of weights. Both replicates and identifiers are provided as part of the ECLS-K data file.

Design effects. In the ECLS-K, a large number of data items were collected from students, parents, teachers, and schools. Each item has its own design effect that can be estimated from the survey data. The median child-level design effect is 4.7 for fall-kindergarten (compared with 2.2 for the National Education Longitudinal Study of 1988 base year student questionnaire data) and 4.1 for spring-kindergarten (compared with 3.4 for the NELS:88 first follow up). The size of the ECLS-K design effects is largely a function of the number of children sampled per school. With about 20 children sampled per school, an intraclass correlation of 0.2 might result in a design effect of about 5. The median design effect is 3.4 for the panel of students common to both fall- and spring-kindergarten, and the lower median design effect is due to the smaller cluster size in the panel. The ECLS-K design effects are slightly higher than the average of 3.8 that was anticipated during the design phase of the study, both for estimates for proportions and for score estimates.

The median teacher-level design effect is 2.5 for both fall- and spring-kindergarten. These are lower than the child-level design effects because the number of responding teachers per school is relatively small. The design effect for teachers is largely a result of selecting a sample using the most effective design for child-level statistics.

The median school-level design effect is 1.6.

A multilevel analysis was carried out to estimate components of variance in fall- and spring-kindergarten cognitive scores associated with the: (1) student level, (2) school level, (3) team leader, and (4) individual test administrator. This secondary analysis was motivated by Westat's earlier finding of larger-than-expected design effects. In addition, the impact on the above sources of variance of the SES indicator (parent's education) was also estimated. It was expected that much of the clustering of students within neighborhood schools (hence higher design effects) could be explained by SES.

Nonsampling Error

During the survey design phase, focus groups and cognitive laboratory interviews were conducted for the purpose of assessing respondent knowledge topics, comprehension of questions and terms, and the sensitivity of items. The design phase also entailed testing for the CAPI instrument and a field test that evaluated the implementation of the survey.

Another potential source of nonsampling error is respondent bias that occurs when respondents systematically misreport (intentionally or unintentionally) information in a study. One potential source of respondent bias in this survey is social desirability bias. If there are no systematic differences among specific groups under study in their tendency to give socially desirable responses, then comparisons of the different groups will accurately reflect differences among the groups. An associated error occurs when respondents give unduly positive assessments about those close to them. For example, parents may give more positive assessments of their children's school experiences than might be obtained from school records or from the teachers.

Response bias may also potentially be introduced in the responses of the teachers about each individual student. Each teacher filled out a survey for each of the sampled children they taught in which they answered questions on the child's socio-emotional development. Since the survey was conducted in the fall it is possible that the teachers did not have adequate time to observe the children, and thus some of the responses may be influenced by the expectations of the teacher based on which groups (e.g., sex, race, linguistic, disability) the children belonged to. In order to minimize bias, all items were subjected to multiple cognitive interviews, field tests, and actual teachers were involved in the design of the cognitive assessment battery and questionnaires. NCES also followed the criteria recommended in a working paper on the accuracy of teacher judgments of students' academic performances (*How Accurate Are Teacher Judgments of Students' Academic Performance?* NCES Working Paper 96-08).

Respondent bias may be present in ECLS-K as in any survey. It is not possible to state precisely how such bias may affect the results. NCES has tried to minimize some of these biases by conducting one-on-one, untimed assessments, and by asking some of the same questions about the sampled child of both teachers and parents.

Coverage error. By designing the ECLS-K child assessment to be both individually administered and untimed, both coverage error and bias were reduced. Individual administration decreases problems associated with group administration such as children slowing down and not staying with the group or simply getting distracted. The advantage of having untimed exams was that the study was able to include most children with learning disabilities, hearing aids, etc. The only children who were excluded from the study were those who were blind, deaf, those whose Individual Education Program (IEP) clearly stated that they were not to be tested, and non-English speaking children who were determined to lack adequate English or Spanish to meaningfully participate in the ECLS-K battery. Exclusion from the direct child assessment did not exclude children from all other parts (e.g., teacher questionnaire, parent interview).

Nonresponse error.

Unit nonresponse. Overall, 944 of the 1,277 schools (74 percent) agreed to participate in the study. More schools participated in the spring of the base year (n=940) than during the fall (n=880), due to the fact that some of the schools that originally declined to participate changed their minds and participated in the spring. Due to the lower than expected cooperation rate for public schools in the fall of the base year, 73 additional public schools were included in the sample as substitutes for schools not participating in the fall. These schools were included in order to meet the target sample sizes for students. Substitute schools are not included in the school response rate calculations.

A nonresponse bias analysis was conducted to determine if substantial bias is introduced due to school nonresponse in ECLS-K. Five different approaches were used to examine the possibility of bias in the ECLS-K sample. First, weighted and unweighted response rates for schools, children, parents, teachers, and school administrators were examined to find large response rate differences by characteristics of schools (e.g., urbanicity, region, school size, percent minority, and grade range) and children (e.g., sex, age, race/ethnicity). Second, estimates based on ECLS-K respondents were compared to estimates based on the full sample. The distributions of schools by school type, urbanicity, region, and the distributions of enrollment by kindergarten type (public versus private), race/ethnicity, urbanicity, region, and eligibility for free and reduced-price lunch were compared for the responding schools and all the schools on the sampling frame. Third, estimates using ECLS-K were compared with estimated from other data sources (e.g., Current Population Survey, National Household Education Survey, Survey of Income and Program Participation). Fourth, estimates using ECLS-K unadjusted weights were compared with estimates using ECLS-K weights adjusted for nonresponse. Large differences in the estimates produced with these two different weights would indicate the potential for bias. Fifth, and last, simulations of nonresponse are being conducted. The results of these analyses are summarized in the ECLS-K User's Manual; however, the findings from these analyses suggest that there is not a bias due to school nonresponse.

The child base-year completion rate was 92 percent; that is, 92 percent of the children were assessed at least once during kindergarten. About 95 percent of the children and 94 percent of the parents who participated in the fall of kindergarten also participated in the spring. Table 1, on the next page, shows how the response rates for those children continued through the spring-1st-grade collection.

Completion rates for the subsample of children included in the Fall-1st-grade collection were 90.3 percent for the children and 88.6 percent for parents. The completion rate for all the children in the spring-1st-grade collection (i.e., including the freshened sample) was 87.2 percent.

BEST COPY AVAILABLE

Table 1. Unit level and overall level weighted response rates for children sampled in kindergarten

Population	Unit level weighted completion rate		
	Base year 1st level	Base year 2nd level	Spring-1st grade
Child assessment	74.2	92.0	88.0
Parent interview	74.2	88.8	84.5
	Overall level weighted response rate		
	Base year 1st level	Base year 2nd level	Spring-1st grade
Child assessment	74.2	68.3	60.1
Parent interview	74.2	62.7	53.0

SOURCE: Tourangeau et al. *ECLS-K Base Year Public-Use Data Files and Electronic Codebook.* Tourangeau et al. *User's Manual for the ECLS-K First Grade Restricted-Use Data Files and Electronic Codebook* (NCES 2001-101).

Measurement error. In addition to the potential clustering effects related to shared parent SES within schools (described in "Design Effects," above), there was a concern in ECLS-K that the individual mode of administration might inject additional and unwanted variance to both the individual and between-school components of variance in the cognitive scores. Since it is more difficult to standardize test administrations when tests are individually administered, this source of variance could contribute to high design effects if the individual assessors differed systematically in their modes of administration. It was found, however, that the component of variance associated with the individual test administration effect was negligible in all three cognitive areas and thus had little or no impact on the design effects.

A potential area for measurement error occurs with the NCATS component of the ECLS-B home visit. The parent-child interaction for this component of the study is videotaped, to be coded later. The process of coding the tapes, however, is not without its problems. The interaction field staff tape must be of high quality to ensure valid coding. For example, field staff should tape the very beginning of the interaction and should not interrupt. The task of coding is further complicated by the coding staff's experience. Like the ECLS-B home visit field staff, ECLS-B NCATS coders do not, for the most part, possess an extensive background in child development. Training the coding staff to reach 90 percent reliability has proven difficult at times, often requiring additional training.

Data Comparability

As a test for nonresponse bias, estimates from ECLS-K are being compared with estimates from other data sources (e.g., Current Population Survey, National Household Education Surveys, Survey of Income and Program Participation).

6. CONTACT INFORMATION

For content information about the ECLS project, contact:

Jerry West
Phone: (202) 502–7335
E-mail: jerry.west@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

ECLS-K, Base Year Public-Use Data File, Kindergarten Class of 1998-99: Data Files and Electronic Code Book; (Child, Teacher, School Files), NCES 2001-029, by K. Tourangeau, J. Burke, T. Le, S. Wan, M. Weant, E. Brown, N. Vaden-Kiernan, E. Rinker, R. Dulaney, K. Ellingsen, B. Barrett, I. Flores-Cervantes, N. Zill, J. Pollack, D. Rosk, S. Atkins-Burnett, S. Meisels, J. Bose, J. West, K. Denton, A. Rathbun, and J. Walston. Washington, DC: 2000.

User's Manual for the ECLS-K First Grade Restricted-Use Data Files and Electronic Codebook, NCES 2001–101, by K. Tourangeau, J. Burke, T. Le, S. Wan, M. Weant, C. Nord, N. Vaden-Kiernan, E. Bissett, R. Dulaney, A. Fields, L. Byrne, I. Flores-Cervantes, J. Fowler, J. Pollack, D. Rock, S. Atkins-Burnett, S. Meisels, J. Bonaki, J. West, K. Denton, A. Rathbun, and J. Walston. Washington, DC: 2001.

Survey Design

Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children, NCES Working Paper 96–18, by S.J. Meisels, S. Atkins-Burnett, and J. Nicholson. Washington, DC: 1996.

A Birth Cohort Study: Conceptual and Design Considerations and Rationale, NCES Working Paper 1999–01, by K. Moore, J. Manlove, K. Richter, T. Halle, S. Le Menestrel, M. Zaslow, A.D. Greene, C. Mariner, A. Romano, and L. Bridges. Washington, DC: 1999.

Formulating a Design for the ECLS: A Review of Longitudinal Studies, NCES Working Paper 97–24, by P.J. Green, L.A. Hoogstra, S.J. Ingels, H.N. Greene, and P.K. Marnell. Washington, DC: 1997.

How Accurate Are Teacher Judgments of Students' Academic Performance? NCES Working Paper 96–08, by N.E. Perry and S.J. Meisels. Washington, DC: 1996.

Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs, NCES Working Paper 97–36, by J.M. Love, A. Meckstroth, and S. Sprachman. Washington, DC: 1997.

A Picture of Young Children's Development: Adapting Assessment Tools for a National Birth Cohort Study, by J. West, L.M. Reaney, and K.L. Denton. Paper presented at the biennial meeting of the Society for Research in Child Development, April 20, 2001, Minneapolis, MN.

Chapter 2: Common Core of Data (CCD)

SURVEY OF THE UNIVERSE OF ELEMENTARY AND SECONDARY SCHOOLS

CCD collects data through these major components:

- Public School Universe Survey
- Public Education Agency Universe Survey
- State Nonfiscal Survey
- National Public Education Finance Survey
- School District Finance Survey
- Early Estimates Survey

1. OVERVIEW

The Common Core of Data (CCD) is NCES' primary database on public elementary and secondary education in the United States. Every year CCD collects information from the universe of state education agencies (SEAs) on all public elementary and secondary schools and education agencies in the United States. CCD provides descriptive data about staff and students at the school, school district, and state levels. Information about revenues and expenditures is collected at the school district and state levels. Some of CCD's component surveys date back to the 1930s. The integrated CCD was first implemented in 1987–88.

Purpose

To provide basic statistical information on all children in this country receiving a public education from prekindergarten through 12th grade and information on the public funds collected and expended for providing public elementary and secondary education. The specific objectives of CCD are: (1) to provide an official listing of public elementary and secondary schools and education agencies in the nation which can be used to select samples for other NCES surveys, and (2) to provide basic information and descriptive statistics on public elementary and secondary schools and schooling.

Components

There are four major components to CCD: the Public School Universe Survey, the Public Education Agency Universe Study, the State Nonfiscal Survey, and the national Public Education Financial Survey. There are also two other surveys: a separate survey that captures early estimates of key items collected in the component surveys (the Early Estimates Survey) and a Census Bureau financial survey that is cross-referenced to CCD (the School District Finance Survey). These surveys are completed by appointed CCD Coordinators in each of the state education agencies for the 50 states, the District of Columbia, the Bureau of Indian Affairs schools, the Department of Defense Dependents Schools, and 5 outlying areas (American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, Puerto Rico, and the Virgin Islands).

Public School Universe Survey. This survey collects information on all of the nearly 91,000 public elementary and secondary schools in the United States. Data include the school's mailing address, telephone number, operating status, locale (ranging from large central city to rural), and type ("regular" or focused on a special area such as vocational education). The survey also collects student enrollment (membership) for every grade taught in the school; number of students in each of five racial/ethnic groups; number of students eligible for free lunch programs; and number of classroom teachers (reported as full-time equivalents). Beginning in 1998–99, several variables were added: location address (if different from mailing); Title I, magnet, and charter school status; number eligible for reduced price lunch programs; migrant students enrolled previous year; and breakout of enrollment by race and sex within grade.

BEST COPY AVAILABLE

Public Education Agency Universe Survey. This survey serves as a directory of basic information on more than 16,000 public education agencies. It collects the agency's mailing address, telephone number, county location, metropolitan status, and type of agency. The survey includes for the *current* year the total number of students enrolled (membership) in grades prekindergarten through 12; number of ungraded students; number of students with Individual Education Programs (IEPs); and number of instructional, support, and administrative staff. It includes for the *previous* year the number of high school graduates, other completers, and grade 7–12 dropouts. Dropout data were first collected in the 1992–93 CCD, reflecting dropouts for the 1991–92 school year. Items that were added in 1998–99 include location address, migrant students provided services during the previous summer, limited English proficiency (LEP) students provided services, and the number of diploma recipients and other high school completers by race and sex.

State Nonfiscal Survey. This survey collects information on all students and staff aggregated to the state level, including number of students by grade level; counts of full-time equivalent staff; and high school completers by race/ethnicity. Data on student enrollment and staffing are for the *current* school year. Data on high school completers and dropouts are for the *previous* year.

National Public Education Financial Survey (NPEFS). This survey collects detailed finance data at the state level, including average daily attendance, school district revenues by source (local, state, federal), and expenditures by function (instruction, support services, and noninstruction) and object (salaries, supplies, etc.). It also reports capital outlay and debt service expenditures.

Early Estimates Survey. This survey collects numbers of students enrolled in public elementary and secondary schools, high school graduates, and teachers, as well as total revenues and expenditures for the operation of public elementary and secondary schools. The survey is designed to allow NCES to report key state-level statistics during the school year to which they apply—compared to 1–2 years later for the other CCD surveys. All *Early Estimates* data are subject to revision.

School District Finance Survey. This survey collects detailed data by school district, including revenues by source, expenditures by function and subfunction, and enrollment. These data are collected through the Bureau of Census' F-33, Annual Survey of Local Government Finances. Data were collected from all districts in the decennial census year (e.g., 1990) and years ending in 2 and 7, and from a large sample in remaining years. Beginning with fiscal year 1995, this is a census. The F-33 data goes back to fiscal year 1980; NCES began to substantially support the survey beginning with the FY 92 collection.

Periodicity

Annual. Some of the component surveys were initiated during the 1930s. CCD, in its integrated form, was introduced in 1986–87.

2. USES OF DATA

CCD collects three categories of information: (1) general descriptive information on schools and school districts, including name, address, phone number, and type of locale; (2) data on students and staff, including demographic characteristics (e.g., race/ethnicity); and (3) fiscal data covering revenues and current expenditures. The datasets within CCD can be used separately or jointly to provide information on many topics related to education. The ease of linking CCD data with other datasets makes CCD an even more valuable resource.

CCD is not only a source of data for demonstrating relationships between different school, district, and state characteristics, but it also provides a historical record of schools or agencies of interest. This information can shed light on how and why education in the United States is changing. The types of schools or districts that have changed the most with respect to a measured characteristic (e.g., proportion of Hispanic students) can be identified, and reasons for these changes can be independently investigated. Similarly, the impacts of state and local education policies and practices can be assessed through an examination of changes in school and district characteristics. For example, districts that have shown substantial improvement in their racial balance or interracial exposure indices can be identified. The policies and practices employed by these districts can then be examined. By identifying the presence of significant changes and where these changes are occurring, CCD data can help policymakers and practitioners better target their efforts and help researchers develop more sharply focused hypotheses for investigating key education issues.

BEST COPY AVAILABLE

3. KEY CONCEPTS

The concepts described below pertain to the levels of data collection (school, agency, state) in CCD. For a comprehensive list of CCD terms and definitions, refer to the glossaries in CCD reports (e.g., *Key Statistics*) and technical user guides available on the Internet and CD-ROM.

Public Education Agency. An agency with administrative responsibility for providing instruction or specialized services to one or more elementary or secondary schools. Most of these agencies are *regular school districts* (also known as local education agencies or LEAs), which are locally administered and directly responsible for educating children. Other agencies include *supervisory unions* (providing administrative systems for smaller regular districts with which they are associated); *regional education service agencies* (offering research, data processing, special education or vocational program management, and other services to a number of client school districts); *state-operated school districts* (e.g., for the deaf and blind); *federally-operated school districts* (e.g., operated by the Bureau of Indian Affairs); and *other agencies* not meeting the definitions of the preceding categories (e.g., operated by a Department of Corrections).

Public Elementary/Secondary School. An institution that is linked with an *education agency*, serves students, and has an administrator. It is possible for more than one CCD-defined school to exist at a single location (e.g., an elementary and secondary school sharing a building, each with its own principal). One school may also spread across several locations (e.g., a multiple "store front" learning center managed by a single administrator).

CCD classifies schools by type. *Regular schools* provide instruction leading ultimately toward a standard high school diploma; they may also offer a range of specialized services. *Special education* and *vocational schools* have the provision of specialized services as their primary purpose. *Other alternative schools* focus on an instructional area not covered by the first three types (e.g., developing basic language and numeracy skills of adolescents at risk of dropping out of school).

Some schools do not report any students in *membership* (i.e., enrolled on the official CCD reporting day of October 1). This occurs when students are enrolled in more than one school but are reported for only one. For example, students whose instruction is divided between a regular and a vocational school may be reported only in membership for the regular school. In other cases, a school may send the students for which it is responsible to another school for their education—a situation most likely in a small community that does not have sufficient students to warrant keeping a school open every year.

4. SURVEY DESIGN

Target Population

All public elementary and secondary schools (nearly 91,000), all LEAs (more than 16,000) and SEAs throughout the United States, including the District of Columbia, the overseas Department of Defense Dependents Schools, and five outlying areas.

Sample Design

CCD collects information from the universe of state-level education agencies.

Data Collection and Processing

CCD data are voluntarily obtained from state administrative records of information collected and edited by the SEA during its regular reporting cycle for the state.

Reference dates. Most data for the nonfiscal surveys are collected for a particular school year (September through August). The official reference date is October 1 or the closest school day to October 1. Special education, free-lunch eligible, and racial/ethnic counts may be taken on December 1 or the closest school day to that date. Student and teacher data are reported for the *current* school year, whereas data for high school graduates, other completers, and dropouts reflect the *previous* year. Fiscal data are for the previous fiscal year, thus FY 98 represents the 1997–98 school year.

Data collection. Survey instruments are usually distributed to the states in January. A State CCD Coordinator, appointed by the Chief State School Officer, is responsible for overseeing the completion of the surveys (the Coordinator for the fiscal surveys is often a different person than for the nonfiscal surveys). To assure comparable data across states, NCES provides the CCD Coordinator with a set of standard critical definitions for all survey items. In addition, data conferences and training sessions are held at least yearly. The state's data plan identifies any definitional differences between the state's recordkeeping and CCD's collection, and any adjustments made by the state to achieve comparability. Counts across CCD surveys may not be identical, but differences should be consistent and the state is asked to describe the reason for the discrepancy.

NCES provides the state with general information collected during the previous survey on each district and school (e.g., name, address, phone number, locale code, and type of school/district). This information must be verified as correct by the CCD Coordinator or recoded with the correct information. The Coordinator must also assign appropriate identification codes to new schools and agencies, and update the operational status codes for schools and agencies that have closed.

CCD data are compiled into prescribed formats and submitted. Nonfiscal data are submitted via diskette or the Internet. Fiscal data are submitted via the web, Internet, diskette, or paper. CCD requests that the data be submitted by March 15 (or the Monday following March 15 if March 15 occurs on a weekend); the CCD nonfiscal closing date to submit the previous year's data is October 1. For fiscal data, the closing date for the current survey year collection is the Tuesday following Labor Day. Corrections to submitted fiscal data are accepted until October 1, but only corrections that lower a state's current expenditure per pupil are accepted after the "Labor Tuesday" deadline for use in the formula for allocating Title I and other ED funding to state and local school systems.

Editing. Completed surveys undergo comprehensive editing by NCES and the states. Where data are determined to be inconsistent, missing, or out of range, NCES contacts the SEAs for verification. States are given the edit software that NCES uses to review their data. They are also asked to confirm prepared summaries of the collected information. At this time, the states may revise data collected in the previous survey cycle. NCES examines the data from the 120 largest school districts on a record-by-record basis, setting up fail-safe edit checks to catch unexplained anomalies. In addition, records are processed through a post-edit to replace blanks and nonmeaningful zeroes with meaningful responses. After editing, final adjustments for missing data are performed.

Early Estimates Survey. The State Coordinators receive survey forms in October and are requested to return them as soon as possible by mail or fax. Coordinators who do not respond by late November are contacted by telephone. All data are checked for reasonableness against prior years' reports, and follow-up calls are made to resolve any questions. When states do not supply a count or estimate, NCES estimates a value. State-supplied estimates that indicate a 10 percent increase or decrease greater than the national average is replaced with NCES estimated values. Early estimates represent the best information available midway through the school year and are reported by NCES in the current school year. All early estimates are subject to later revision.

Estimation Methods

NCES estimates missing values to improve data comparability across states. Only state-level data are estimated on a regular basis. Missing values in the Public School and Agency Universe Surveys are generally left as missing, with a few exceptions.

There are two basic estimation methods: imputation and adjustment. ***Imputation*** is performed when the missing value for a data item is not reported at all, indicating that subtotals and totals containing the category are underreported. Imputation assigns a value to the missing item, and the subtotals and totals containing this item are increased by the amount of the imputation. ***Adjustment*** corrects a situation in which a value reported for one item contains a value for one or more additional items not reported elsewhere. The original value is reduced by an appropriate amount, which is distributed to the items missing a value. All totals and subtotals are then recalculated. If it is not possible to impute or adjust for a missing value, the item remains blank and is counted as "missing."

Every cell in the data file has a companion cell with a flag indicating whether the data contents were reported by the state (R) or placed there by NCES using one of several methodologies: adjustment (A); imputation based on the prior year's data (P); imputation based on a method other than the prior year's data (I); totaling based on the sum of internal or external detail (T); or combining with data provided elsewhere by the state (C).

Estimating state-level nonfiscal data. NCES imputes and adjusts some reported values for student and staff counts at the state level (including the District of Columbia). Imputations for prekindergarten students are performed first, followed by staff imputations and then other adjustments. No imputations or adjustments are made to racial/ethnic data.

Estimating state-level fiscal data. NCES also imputes and adjusts revenue and expenditure data. The federal standard, defined in *Financial Accounting for Local and State School Systems, 1990,* is used in the adjustments to distributed expenditure and revenue data. Adjustments are also used to distribute direct state support expenditures to specific objects and functions. In come cases, local revenues from student activities and food services are imputed.

Early Estimates Survey. NCES imputes values for *Early Estimates* data when the states themselves do not provide preliminary counts or their own estimates of counts.

Future Plans

Because it is an ongoing annual survey, CCD engages in continuous planning with its data users and providers. Changes are likely in 2004 due to the newly revised NCES Financial Accounting Handbook and new reporting implementation guidelines set by the Government Accounting Standards Board. The 2004 CCD will also incorporate tabulation guidelines for the newly approved racial and ethnic definitions.

NCES has contracted with the Census Bureau to produce a standardized district finance file and file documentation (meeting formal NCES requirements) for fiscal years 1990 to 1998. This work is still in progress.

5. DATA QUALITY AND COMPARABILITY

The data in CCD are obtained from the universe of SEAs, which are provided with a common set of definitions for all data items requested. In addition, NCES provides crosswalk software which converts a state's existing accounting reports to the federal standard, as indicated in *Financial Accounting for Local and State School Systems, 1990.* This ensures the most comparable and comprehensive information possible across states. As with any survey, however, there are possible sources of error, as described below.

Sampling Error

Because CCD is a universe survey, its data are not subject to sampling errors.

Nonsampling Error

Coverage error. A recent report, *Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary Education Agency Universe Survey* (NCES 97–505), found that overall coverage in the Agency Universe Survey was 96.2 percent (in a comparison to state education directories). "Regular" agencies—those traditionally responsible for providing public education—had almost total coverage in the 1994–95 survey. Most coverage discrepancies were attributed to nontraditional agencies that provide special education, vocational education, and other services.

Nonresponse error.
Unit nonresponse. The unit of response in CCD is the state education agency. Under current NCES standards, the regular components of CCD are likely to receive at least partial information from every state, resulting in a 100 percent unit response rate.

Item nonresponse. Any data item missing for one school district is generally missing for other districts in the same state. The following items have higher than normal nonresponse: free-lunch-eligible students by school; nonregular agencies; and dropouts. Some states assign all ungraded students to one grade and therefore do not report any ungraded students.

Several items have shown marked improvement in response during recent years. Student enrollment was only reported for 80 percent of the districts in 1986–87, but is now available for about 100 percent. Reports of student race/ethnicity at the school level increased from 63 percent in 1987–88 (when first requested) to nearly 100 percent today.

Measurement error. Measurement error typically results from varying interpretations of NCES definitions, differing recordkeeping systems in the states, and failures to distinguish between zero, missing, and inapplicable in the reporting of data. NCES attempts to minimize these errors by working closely with the state CCD Coordinators.

Definitional differences. Although states follow a common set of definitions in their CCD reports, the differences in how states organize education lead to some limitations in the reporting of data, particularly regarding dropouts. CCD definitions appear to be less problematic for NPEFS Coordinators, although data on average daily attendance in NPEFS are not comparable across states. States provide figures for average daily attendance in accordance with state law; NCES provides a definition for states to use in the absence of state law. Because of this lack of comparability, student membership counts from the State Nonfiscal Survey are used as the official state counts.

Because not all states follow the CCD *dropout* definition and reporting specifications, dropout counts cannot be compared accurately across states. For states that do not

BEST COPY AVAILABLE

comply with the CCD definition, the dropout count is blanked out in the database and considered missing. Currently, there is considerable variation across local, state, and federal data collections on how to define dropouts. CCD's definition differs from that in other data sources, including the High School and Beyond Study, the National Education Longitudinal Study of 1988, and the Current Population Survey (CPS, conducted by the Bureau of the Census). Although the collection of dropout information in CCD was designed to be consistent with procedures for the CPS, differences remain. CCD dropout data are obtained from state administrative records (whereas CPS obtains this information from a household survey). CCD includes dropouts in grades 7 through 12 (whereas CPS includes only grades 10 through 12).

States also vary in the kinds of high school completion credentials on which they collect data. Some issue a single diploma regardless of the student's course of study. Others award a range of different credentials depending upon whether the student completed the regular curriculum or addressed some other individualized set of education goals. Unreported information is shown as missing in CCD data files and published tables unless it is possible to impute or adjust a value (see section 4, Estimation Methods).

Changes in state reporting practices. Basic characteristics of a school or district do not change frequently. However, a minor change in local or statewide reporting practices (such as two or three Coordinators instructing schools to review *all* of their general information) can have a large impact on the reliability and validity of CCD items. In 1990–91, a significant proportion (7 percent) of schools, primarily in three states, reported a change in locale code from the prior survey. While this undoubtedly provided better information on school locales in these states, data became less comparable across years. Such changes are rare, however, and tend to be clustered by state and year.

Data Comparability

Most CCD items can be used to assess changes over time by state, district, and school. However, checks of the prevalence and patterns of nonresponse should be performed to assess the feasibility of any analysis. There may also be discontinuities in the data resulting from the introduction of new survey items, changes in state reporting practices, etc., and there may be inconsistencies across reporting levels in the numbers for the same data element (e.g., number of students).

Content changes. As new items are added to CCD, NCES encourages the states to incorporate into their own survey systems the items they do not already collect so that these data will be available in future rounds of CCD. Over time, this has resulted in fewer missing data cells in each state's response, thus reducing the need to impute data. Users should keep in mind, however, that while the restructuring of data collection systems can produce more complete and valid data, it can also make data less comparable over time. For example, prior to fiscal year 1989, public revenues were aggregated into four categories and expenditures into three functions. Because these broad categories did not provide policymakers with sufficient detail to understand changes in the fiscal conditions of states, the survey was expanded in 1990 to collect detailed data on all public revenues and expenditures within states for regular prekindergarten to grade 12 education.

Comparisons within CCD. A major goal of CCD is to provide comparable information across all surveys. The surveys are designed so that the schools in the Public School Universe are those reflected in the Public Agency Universe, and so that the data from these universes are reflected in the state aggregate surveys. While counts may not always be equal across reporting levels or even within the same level, differences should be consistent and explainable. For example, counts of students by race/ethnicity in the Public School Universe may not always be comparable to student counts by grade because these counts may be taken at different times.

For the most part, the total number of students in a regular district is close to the aggregated number of students in all of the district's schools. Since 1990, there has typically been agreement between these counts in at least 85 percent of the districts. Membership numbers in the Public School and Agency Universes may legitimately differ if: (1) there are students served by the district but not accounted to any school (e.g., hospitalized or homebound students), or (2) there are schools operated by the state Board of Education rather than by a local agency. To avoid confusion, NCES publishes the numbers of students and staff from the State Nonfiscal Survey as the official counts for each state.

Teacher counts may also vary across reporting levels. Teachers are reported in terms of full-time equivalency (FTE), rounded to the nearest tenth, in the Public School Universe. FTE teacher counts are rounded to the nearest whole number in the State Nonfiscal Survey.

Comparisons with the Early Estimates Survey. Early estimates are reported midway through the school year and do not undergo the verification and editing procedures required for the other CCD surveys. All early estimates are subject to revision once the data from the other CCD surveys are verified and adjustments completed. Numbers for a given data item in *Early Estimates* publications are likely to differ somewhat from numbers for that same data item reported in later NCES publications. Nevertheless, comparisons of estimated change from 1994–95 to 1995–96 (as reported in the *Early Estimates Survey*) and actual change (as reported in the regular CCD surveys) reveal differences of less than one percentage point for membership, high school graduates, current expenditures, and revenues. Of the five changes compared, only teachers showed a larger discrepancy, with *Early Estimates* projecting an increase of 1.5 percent and CCD reporting an actual decrease of 0.1 percent between the two surveys. For nearly all states, the early estimates were within 10 percent of the final reported CCD counts for these items.

6. CONTACT INFORMATION

For content information on CCD, contact the following individuals:

Public School Universe and Public Education Agency Universe:
John Sietsema
Phone: (202) 502–7425
E-mail: john.sietsema@ed.gov

State Nonfiscal Report:
Beth Young
Phone: (202) 502-7480
E-mail: beth.young@ed.gov

National Public Education Finance Survey, and School District Finance Survey:
Frank Johnson
Phone: (202) 502–7362
E-mail: frank.johnson@ed.gov

Early Estimates Survey:
Lena McDowell
Phone: (202) 502–7396
E-mail: lena.mcdowell@ed.gov

Frank Johnson
Phone: (202) 502–7362
E-mail: frank.johnson@ed.gov

Mailing Address for All Contacts:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

Data Quality and Comparability

Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary Education Agency Universe Survey, NCES 97–505, by S. Owens and J. Bose. Washington, DC: 1997.

Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary School Universe Survey, NCES Working Paper 2000–12, by T. Hamann. Washington, DC: 2000.

Customer Service Survey: Common Core of Data Coordinators, NCES Working Paper 97–15, by L. Hoffman. Washington, DC: 1997.

Disparities in Public School District Spending 1989–90: A Multivariate, Student-weighted Analysis, Adjusted for Differences in Geographic Cost of Living and Student Need, NCES 95–300, by T.B. Parrish, C.S. Matsumoto, and W.J. Fowler. Washington, DC: 1995.

Survey Design

Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle, NCES Working Paper 1999–03, by T.A. Hamann. Washington, DC: 1999.

Chapter 3: Private School Universe Survey (PSS)

1. OVERVIEW

In recognition of the importance of private education, NCES has made the collection of data on private elementary and secondary schools a priority. In 1988, NCES introduced a proposal to develop a Private School Data Collection System that would improve on the irregular collection of private school information dating back to 1890. Since 1989, the U.S. Bureau of the Census has conducted the biennial Private School Universe Survey (PSS) for NCES. PSS collects information comparable to that collected on public schools in the Common Core of Data (CCD—see chapter 2). PSS data are complemented by more in-depth information collected in the private school sample surveys that are part of the Schools and Staffing Survey (SASS—see chapter 4). The next PSS data collection will take place during the 2003–04 school year. The next SASS is planned for the 2003–04 school year.

BIENNIAL SURVEY OF THE UNIVERSE OF PRIVATE SCHOOLS

PSS collects data on:
- Student enrollment
- Teaching staff
- High school graduates
- School religious affiliation

Purpose

To (1) build an accurate and complete universe of private schools to serve as a sampling frame for NCES surveys of private schools, and (2) generate biennial data on the total number of private schools, teachers, and students.

Components

PSS consists of a single survey that is completed by administrative personnel in private schools. An early estimates survey designed to allow early reporting of key statistics was discontinued after the 1992–93 school year.

Private School Universe Survey. This survey collects data on private elementary and secondary schools, including: religious orientation, level of school, size of school, length of school year, length of school day, total enrollment (K–12), race/ethnicity of students, number of high school graduates, number of teachers employed, program emphasis, and existence and type of kindergarten program.

Periodicity

Biennial. The next PSS will be administered in 2003–04 and then every 2 years thereafter. Earlier surveys were conducted in 1989–90, 1991–92, 1993–94, 1995–96, 1997–98, 1999–2000, and 2001–02.

2. USES OF DATA

PSS produces private school data similar to that for public schools in CCD. Profiles of private education providers can be developed from PSS data to address a variety of policy- and research-relevant issues, including the growth of religiously-affiliated schools, the number of private high school graduates, the length of the school year for various private schools, and the number of private school students and teachers.

NCES uses an indirect estimate approach as an alternative to the current procedures for the production of state estimates of the number of private schools in the nation and the associated numbers of students, teachers, and graduates. (See *Indirect State-level Estimation for the Private School Survey*, NCES 1999–351).

3. KEY CONCEPTS

Some key concepts related to PSS are described below.

Private School. A school that is not supported primarily by public funds. It must provide instruction for one or more of grades K through 12 (or comparable ungraded levels), and have one or more teachers. Organizations or institutions that provide support for home schooling but do not offer classroom instruction for students are not included. Private schools are assigned to one of three major categories and, within each major category, to one of three subcategories:

- *Catholic*: parochial, diocesan, private;
- *Other religious*: affiliated with a conservative Christian school association, affiliated with a national denomination, unaffiliated; and
- *Nonsectarian*: regular program emphasis, special program emphasis, special education.

Schools with kindergarten, but no grade higher than kindergarten, are referred to as ***kindergarten-terminal (K-terminal)*** schools; these schools were first included in the 1995–96 PSS. Schools meeting the pre-1995 definition of a private school (i.e., including any of grades 1 through 12) are referred to as ***traditional*** schools.

Elementary School. A school with one or more of grades K–6 and no grade higher than grade 8. For example, schools with grades K–6, 1–3, or 6–8 are classified as elementary schools.

Secondary School. A school with one or more of grades 7–12 and no grade lower than grade 7. For example, schools with grades 9–12, 7–8, 10–12, or 7–9 are classified as secondary schools.

Combined School. A school with one or more of grades K–6 and one or more of grades 9–12. For example, schools with grades K–12, 6–12, 6–9, or 1–12 are classified as combined schools. Schools in which all students are ungraded (i.e., not classified by standard grade levels) are also classified as combined.

Teacher. Any full-time or part-time teacher whose school reports that his or her assignment is teaching in any of grades K–12.

4. SURVEY DESIGN

Target Population

All private schools in the United States that meet the NCES definition. The PSS universe consists of a diverse population of schools. It includes both schools with a religious orientation (e.g., Catholic, Lutheran, or Jewish) and nonsectarian schools with programs ranging from regular to special emphasis and special education.

Sample Design

NCES uses a dual frame approach for building its private school universe. The primary source of the PSS universe is a *list frame* containing most private schools in the country. The list frame is supplemented by an ***area frame***, which contains additional schools identified during a search of randomly selected geographic areas around the country. The two frames are used together to estimate the population of private schools in the United States.

List frame. In an effort to ensure a complete population list of all private elementary and secondary schools in the United States, NCES updates the list frame every 2 years in preparation for the next PSS administration. This frame, developed over more than a decade, is assembled from lists provided by several sources, including private school associations and state departments of education. The lists from these sources are matched against the most recent PSS universe. Nonmatches are added to the universe as births.

The basis of the current survey's list frame is the previous PSS. In order to expand coverage to include private schools founded since the previous survey, NCES requests lists of schools from the 50 states and the District of Columbia in advance of each survey administration. Requests are made to state education departments, as well as to other departments such as health or recreation. NCES also collects membership lists from about 26 private school associations and religious denominations. Schools on the state and association lists are compared to the base list, and any school not matching a school on the base list is added to the universe list.

Prior to the 1995–96 survey, only schools that included at least one of grades 1–12 were included in PSS (now referred to as *traditional schools*). As of 1995–96, PSS also collects data from schools for which kindergarten is the highest grade (referred to as *K-terminal schools*). NCES also removed from the PSS eligibility criteria the requirements that a school have 160 days in the school year and 4 hours per day conducting classes. The list of K-terminal schools for the 1999–2000 PSS was assembled from state and association lists and information obtained from questionnaires sent to about 5,800 programs identified in the 1997–98 PSS as prekindergarten only.

Area frame. The list frame is supplemented by an *area frame* containing additional private schools identified during a search of telephone books and other sources in randomly selected geographic areas around the country. Each area's list is created from a set of predetermined sources within that area and then matched against the updated list frame universe to identify schools missing from the updated list frame.

The United States is divided into 2,054 primary sampling units (PSUs), each consisting of a single county, independent city, or cluster of geographically contiguous areas. During the first NCES area search for private schools conducted in 1983, eight PSUs with populations greater than 1.7 million were selected with certainty for the private school survey; these same eight PSUs have been retained as certainty PSUs in all PSS administrations. In addition to these certainty PSUs, the area frame consists of two sets of sample PSUs: (1) a 50 percent subsample (overlap) of the area frame sample PSUs from the previous PSS, maintaining a reasonable level of reliability in estimates of change, and (2) a sample of PSUs selected independently from the previous PSS sample (nonoverlap). A minimum of two nonoverlap PSUs are allocated to each of the 16 strata, which are defined as follows: (a) four Census regions (Northeast, Midwest, South, West); (b) metro/nonmetro status (two levels); and (c) whether the PSU's percentage of private school enrollment exceeds the median percentage of private enrollment of the other PSUs in the census region/metro status strata (two levels). Within a stratum, the sample PSUs are selected with probability proportional to the square root of the population in each of the PSUs.

The 1999–2000 area sample included a total of 125 distinct PSUs (sampled geographic areas). Within each of these PSUs, the Census Bureau attempted to find all eligible private schools. A block-by-block listing of all private schools in a sample of PSUs was not attempted. Instead, regional field staff created the frame by using sources such as the yellow pages, local Catholic dioceses, religious institutions, local education agencies, and local government offices. Once the area search lists were constructed, they were matched against the list frame. Schools not matching the list frame were considered part of the area frame.

Due to differences in methodology and definition, the results of the 1993–94 and subsequent area search frames are not strictly comparable to results in earlier years. Prior to 1993, an initial eligibility screening was performed over the telephone for area frame schools before the questionnaire was mailed out. Ineligible schools were declared out of scope at that time, and eligible schools were either interviewed over the telephone or sent a questionnaire. In the 1993–94 PSS, screener questions were added to the survey instrument for the purpose of determining eligibility. Ineligible schools were not eliminated until after the questionnaires were returned. In the 1995–96 PSS, all area frame schools were placed in the telephone follow-up phase of PSS, and ineligible schools were again eliminated based on responses to screener questions.

Since 1995–96, schools are no longer required to have 160 days in the school year or to conduct classes for at least 4 hours per day to be included. The combination of these changes resulted in an increased number of schools surveyed in the last two surveys.

Data Collection and Processing

The data collection phase consists of (1) a mailout/mailback stage and (2) a telephone follow-up stage. The U.S. Bureau of the Census is the collection agent.

Reference dates. The official reference date for reporting PSS information is October 1.

Data collection. In October of the survey year, the Census Bureau mails PSS questionnaires to the private schools. (Data collection for the 1999–2000 PSS coincided with the data collection phase of the private school component of the 1999–2000 SASS: the private schools selected for SASS were excluded from PSS, and the schools selected for SASS received a SASS private school questionnaire only, while the remaining private schools were sent a PSS questionnaire. The PSS questionnaire used the same wording as the SASS questionnaire, but contained only a subset of the SASS questionnaire items. After data collection, the data for the SASS cases were merged into the PSS universe.) If no response is received within a month, a second questionnaire is mailed. Reminder postcards are sent 1 week after each questionnaire mailout. Three to 4 months after the initial mailout, the Census Bureau begins telephone follow up of schools that have not responded to either mailout; the schools from the area frame operation are added at this time. Interviewing takes place at the Census Bureau's computer-assisted telephone interviewing (CATI) facilities. For schools that cannot be contacted by telephone, additional follow up is conducted in the Census Bureau's Regional Offices.

The 1999–2000 PSS return rate (i.e., the total number of returns—interviews, noninterviews, and out-of-scopes—divided by the total number of schools in the Private School Universe) was 40 percent at the end of the first mailout and 62 percent at the end of the second mailout. Follow-up efforts achieved a final unweighted return rate of 100 percent.

Editing. Most of the mailback questionnaires are scanned; those that must be keyed are 100 percent key-verified. For data collected during the telephone follow-up phase, preliminary quality assurance and editing checks take place at the time of the interview. The data collection instrument is designed to alert interviewers to inconsistencies reported by the respondent so that any necessary corrections can be made at this time. Data from the CATI facilities are transmitted to Census headquarters for further processing. All data then undergo extensive editing at the Census Bureau's headquarters. The edits include:

- range checks to eliminate out-of-range entries;
- consistency edits to compare data in different fields for consistency;
- blanking edits to verify that skip patterns on the questionnaire were followed; and
- interview status recodes (ISR), performed prior to the weighting process to assign
- the final interview status to the records (i.e., interview, noninterview, or out-of-scope, as described above).

Estimation Methods

Weighting adjusts the number of schools in the area frame sample up to a fully representative number of schools missing from the list frame, and adjusts the survey data from both the area and list components for school nonresponse. Imputation is used to compensate for item nonresponse.

Weighting. PSS data from the area frame component are weighted to reflect the sampling rates (probability of selection) in the PSUs. Survey data from both the list and area frame components are adjusted for school nonresponse. This represents a departure from procedures used in the 1989 survey, which adjusted for total nonresponse (i.e., school nonresponse) and for partial nonresponse associated with four specific PSS data elements. Since 1991, only one weight has been required, due to a newly developed and complex imputation process used to compensate for item nonresponse. When estimates are produced for schools and other data elements, the same PSS school weight should be used. A brief description of the components comprising the PSS weight follows:

W_i, the PSS weight for all data items for the i^{th} school is:

$$W_i = BW_i \times NR_c$$

where: BW_i is the inverse of the selection probability for school i ($BW_i = 1$ for list frame schools; BW_i = inverse of the PSU probability of selection for area frame schools), and

NR_c is the weighted ratio of the sum of the in-scope schools to the sum of the in-scope responding schools in cell c, using BW_i as the weight.

The cells used in NR_c are school association by school level, by size, by urbanicity for list frame schools; the cells used in NR_c for area frame schools are certainty/noncertainty PSU by school affiliation by school level. If the number of schools in cell c is less than 15 or NR_c is greater than 1.5, then cell c is collapsed. List frame cells for traditional schools were collapsed within enrollment category, urbanicity and grade level. Associations were never collapsed together. List frame cells for k-terminal schools were collapsed within enrollment category and urbanicity before the associations were collapsed. Area frame cells for traditional schools were collapsed within grade level before affiliation cells (Catholic, other religious, nonsectarian) were collapsed. Area frame cells for k-terminal schools were collapsed within affiliation.

Imputation. Since the 1991–92 PSS, imputation has been used to compensate for item nonresponse in records classified as interviews (i.e., required items are completed). All items that are missing data are imputed. The first survey, the 1989–90 PSS, used weighting adjustments for both interviews and noninterviews.

Imputation occurs in two stages. The first stage (internal) imputation uses data from other items for the same school in the current PSS and data from the previous PSS. If an item cannot be imputed during the first stage processing, it is imputed during the second stage. The second stage (donor) process uses a hot-deck imputation methodology that extracts data from the record for a reporting school (donor) similar to the nonrespondent school. All records (donors and nonrespondents) on the file are sorted by variables that describe certain characteristics of the schools, such as school type, affiliation, school level, enrollment, and urbanicity.

For a few items, there are cases where entries are clerically imputed. The data record, sample file record, and the questionnaire are reviewed and an entry consistent with the information from those sources is imputed. This procedure is used when: (1) no suitable donor is found, (2) the computer method produces an imputed entry that is unacceptable, and (3) the nature of the item requires an actual review of the data rather than a computer-generated value.

Recent Changes

Several changes to the questionnaire were introduced in the last few PSS cycles. Three major revisions were made to the *1993–94 PSS*. First, a new design was implemented to facilitate respondent reporting by clearly indicating skip patterns through the use of arrows as well as words and by minimizing the number of questions asked on each page. Second, content on prekindergarten programs was expanded to collect the type of prekindergarten program in addition to the prekindergarten student and teacher counts requested in earlier surveys. Third, data on the racial/ethnic makeup of the school's student body were collected for the first time.

Modifications made to the *1995–96 PSS* included adding nursery and prekindergarten, transitional kindergarten, and transitional first grade enrollment counts to the enrollment item. Questions regarding the length of school day and number of days per week for kindergarten, transitional kindergarten, and transitional first grade were also added. "Early childhood program/day care center" was added as a category for type of school. Items on types of prekindergarten programs and the number of prekindergarten teachers were deleted.

In the *1997–98 PSS*, the following items were added to the survey instrument: (1) whether or not the school is coeducational (and if yes, the number of male students; if no, whether the school is all female or all male); and (2) whether or not the school has a library or library media center.

There were few changes in the *1999–2000 PSS*. One religious affiliation—Church of God in Christ—was added, and three associations were added—Association of Christian Teachers and Schools, National Coalition of Girls' Schools, and state or regional independent school association. The item that previously collected data on the number of graduates that applied to 2-year or 4-year colleges was changed to collect data on the percentage of graduates who went on to attend three types of schools: 2-year colleges, 4-year colleges, and technical or other specialized schools.

Future Plans

PSS will continue as a biennial survey.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Only the area frame contributes to the standard error in PSS. The list frame component of the standard error is always 0. Estimates of standard errors are computed using half-sample replication.

Because the area frame sample of PSUs is small (125 out of a total of approximately 2,000 eligible PSUs), there is a potential for unstable estimates of standard errors. This is particularly true when the domain of interest is small and there may not be enough information to compute a standard error. Stabilizing the standard error estimate given the level of detail of the PSS estimates would require a much larger PSU sample. The current area frame is designed to produce regional estimates.

Nonsampling Error

Coverage error. Undercoverage is one possible source of nonsampling error. Because PSS uses a dual frame approach, it is possible to estimate the coverage or completeness of PSS. A capture-recapture methodology is used to estimate the number of private schools in the United States and to estimate the coverage of private schools. The coverage rate for schools was equal to 97 percent in the 1999–2000 PSS.

A study evaluating the quality of PSS frame coverage in comparison to the commercial Quality Education Data database of schools is discussed by Hynshik Lee, John

Burke, and Keith Rust in their paper "Evaluating the Coverage of the U.S. National Center for Education Statistics' Public and Private School Frames Using Data from the National Assessment of Educational Progress," published in the *Proceedings of the Second International Conference on Establishment Surveys*.

Nonresponse error.

Unit nonresponse. The unweighted unit response rate for traditional schools in the 1999–2000 PSS was 93.1 percent, and the weighted response rate was 92.7 percent. For K-terminal schools in the 1999–2000 PSS, the unweighted response rate was 98.4 and the weighted response rate was 98.6 percent.

Item nonresponse. For traditional schools, all but three items in the 1999–2000 PSS had unweighted response rates greater than 90 percent. The three lower rates (ranging from 76.1 percent to 82.8 percent) pertained to the percentage of graduates who went to 4-year colleges, 2-year colleges, and technical or other specialized schools. Imputation is used to compensate for item nonresponse.

Measurement error. NCES seeks to minimize measurement error by developing survey content in consultation with representatives of private school associations, reviewing extensively the questionnaire and instructions before distribution, requiring that the data that are not scanned are 100 percent key-verified, and processing the survey data through an extensive series of edits to verify accuracy and consistency.

Intersurvey Consistency in NCES Private School Surveys

PSS and the private school component of SASS were fielded in the same school year for the first time in 1993–94. Even though these two surveys measure some of the same variables (schools, teachers, and students), the 1993–94 results were not in agreement due to sampling and other errors. PSS results are likely to be the more accurate since PSS serves as the sampling frame for the SASS private school component (a sample of around 3,000 schools). Special methodological studies of these two surveys have been done, including empirical results of attempts to ensure that the 1993–94 PSS numbers of schools, teachers, and students was the same as the 1993–94 SASS numbers of private schools, private school teachers, and private school students—see *Intersurvey Consistency in NCES Private School Surveys* (NCES Working Paper 95–16) and *Intersurvey Consistency in NCES Private School Surveys for 1993–94* (NCES Working Paper 96–27).

Data Comparability

While changes to survey design and content generally result in improved data quality, they also impact the comparability of data over time. Recent changes to PSS and the comparability of PSS data both within PSS itself and with other data sources are discussed below.

Design change. Changes in the survey design of the 1995–96 PSS resulted in an increased number of private schools in the survey population. First, seven new association lists were obtained, adding 512 new schools to the list frame. In previous years, the area frame was relied upon to include these schools. Second, the area search results were not strictly comparable to those in previous years due to procedural differences. The 1995–96 PSS was the first survey to verify the control of schools marked as public in the screener item. Final determination of school control was based on a review of the school's name and other identifying information. As a result, several schools marked as public but obviously private were added back into PSS. They were counted as interviews if the required data were provided or as noninterviews if the required data were missing. Third, the eligibility criteria for PSS were changed to no longer require schools to have 160 days in the school year or to conduct classes for at least 4 hours per day. Fourth, the PSS definition of a school was expanded to include programs where kindergarten is the highest grade (K-terminal schools). Additional lists of programs which might have a kindergarten were requested from nontraditional sources, and the area search was expanded to search for programs with a kindergarten. Some schools meeting the traditional PSS definition of a school (any of grades 1–12 or comparable ungraded levels) were discovered on these lists. When added to PSS, these schools also increased the estimates of traditional schools.

Note that even when the population of schools is about the same from one survey to the next, it may represent a different set of schools. For example, the number of schools was around 27,000 in both 1997–98 and 1999–2000, although about 1,700 schools were added to the PSS universe in 1999–2000. This suggests that a nearly equal number of schools dropped out of the universe between 1997–98 and 1999–2000.

Questionnaire changes. Several modifications have been made to both the format and content of the PSS questionnaire since 1991–92. A number of items were added (including race/ethnicity of students), and some items were deleted or modified.

Comparisons within PSS. Comparisons of the 1999–2000 PSS estimates with those from previous surveys show no significant change in the estimates for the number of private schools; however, the estimates do indicate an increase in the estimate for the number of teachers and number of private school students.

Comparisons with the Current Population Survey. A comparison of the PSS estimates of K–12 students enrolled in all private schools in the 1999–2000 school year with the household survey estimate from the October 1999 Supplement of the Current Population Survey (CPS) shows that the PSS estimate of 5,254,485 is lower than the CPS estimate of 5,532,000; the 95 percent confidence interval on the CPS estimate ranges from 5,314,000 to 5,750,000. The 1997–98 PSS estimate was larger than the CPS estimate (5,179,180 to 4,883,000, respectively) and fell above the upper 95 percent confidence interval on the CPS estimate. The 1995–96 PSS estimates of K–12 students was within the CPS confidence interval (5,146,753 to 5,324,000, respectively). Prior to 1995–96, the PSS estimate did not include kindergarten enrollment from K-terminal schools, whereas the CPS has always included kindergarten enrollment from K-terminal schools.

Comparisons with National Catholic Educational Association data. Comparisons of the PSS estimate for Catholic schools with the National Catholic Educational Association (NCEA) data for the 1999–2000 school year show a similarity in school counts but a difference in the student counts. Beginning in the 1997–98 school year, the NCEA computed FTE teacher counts giving each part-time teacher a weight of 0.333. Therefore, the FTE teacher counts are not strictly comparable between PSS and NCEA. The survey methodologies used by NCES and NCEA are quite different; NCES surveys private schools directly while NCEA surveys archdiocesan and diocesan offices of education and some state Catholic conferences. The NCEA 1999–2000 school year count of 8,144 schools was within the 95 percent confidence interval of the 1999–2000 PSS estimate for Catholic schools (ranging from 8,054 to 8,150). However, the NCEA K–12 student count of 2,500,416 was lower than the 95 percent confidence interval of the 1999–2000 PSS estimate for Catholic students (ranging from 2,501,659 to 2,520,422). Both the NCEA teacher count of 157,134 and the PSS estimate of 149,600 include part-time and full-time teachers in the computation of full-time equivalents (the 95 percent confidence interval of the PSS estimate ranges from 149,188 to 150,012).

NCES publication criteria for PSS. NCES criteria for the publication of an estimate are dependent on the type of survey—sample or universe. To publish an estimate for a sample survey, at least 30 cases must be used in developing the estimate. For a universe survey, a minimum of three cases must be used. PSS includes both types of surveys: (1) a sample survey of PSUs (area frame) which collects data on schools not on the list frame (the number of PSUs changes for each administration), and (2) a complete census of schools belonging to the list frame. NCES has established a rule that published PSS estimates must be based on at least 15 schools. If the estimate satisfies this criterion and the coefficient of variation (standard error/estimate) is greater than 25 percent, then the estimate is identified as having a large coefficient of variation and the reader is referred to a table of standard errors.

6. CONTACT INFORMATION

For content information on PSS, contact:

Stephen Broughman
Phone: (202) 502–7315
E-mail: stephen.broughman@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

Methodology discussed in Technical Notes.

General

Private School Universe Survey, 1999–2000, NCES 2001–330, by S.P. Broughman and L.A. Colaciello. Washington, DC: 2001.

Private School Universe Survey, 1997–98, NCES 1999–319, by S.P. Broughman and L.A. Colaciello. Washington, DC: 1999.

Private School Universe Survey, 1995–96, NCES 98–229, by S. Broughman and L. Colaciello. Washington, DC: 1998.

Private School Universe Survey, 1993–94, NCES 96–143, by S. Broughman. Washington, DC: 1996.

Private School Universe Survey, 1991–92, NCES 94–350, by S. Broughman, E. Gerald, L.T. Bynum, and K. Stoner. Washington, DC: 1994.

Private School Universe Survey, 1989–90, NCES 93–122, by E. Gerald, M. McMillen, and S. Kaufman. Washington, DC: 1992.

Survey Design

Diversity of Private Schools, NCES 92–082, by M. McMillen and P. Benson. Washington, DC: 1992.

Intersurvey Consistency in NCES Private School Surveys, NCES Working Paper 95–16, by F. Scheuren and B. Li. Washington, DC: 1995.

Intersurvey Consistency in NCES Private School Surveys for 1993–94, NCES Working Paper 96–27, by F. Scheuren and B. Li. Washington, DC: 1996.

Data Quality and Comparability

"Evaluating the Coverage of the U.S. National Center for Education Statistics' Public and Private School Frames Using Data from the National Assessment of Educational Progress," *The Second International Conference on Establishment Surveys: Survey Methods for Businesses, Farms, and Institutions* (pp. 89–98), by H. Lee, J. Burke, and K. Rust. Arlington, VA: American Statistical Association. 2000.

Improving the Coverage of Private Elementary-Secondary Schools, NCES Working Paper 96–26, by B.J. Jackson and R.L. Frazier. Washington, DC: 1996.

"Improving the Coverage of Private Elementary-Secondary Schools," in *Selected Papers on the Schools and Staffing Survey: Papers Presented at the 12997 Meeting of the American Statistical Association*, NCES Working Paper 97–41, by B.J. Jackson, N.R. Johnson, and R.L. Frazier. Washington, DC: 1997.

Indirect State-Level Estimation for the Private School Survey, NCES 99–351, by B.D. Causey, L. Bailey, and S. Kaufman. Washington, DC: 1999.

Chapter 4: Schools and Staffing Survey (SASS)

1. OVERVIEW

The Schools and Staffing Survey (SASS) provides data on public and private schools, principals, school districts, and teachers. SASS gathers information about many topics, including various characteristics of elementary and secondary students, some of the professional and paraprofessional staff who serve them, the programs offered by schools, principals' and teachers' perceptions of school climate and problems in their schools, teacher compensation, and district hiring practices. SASS is a unified set of surveys that facilitates comparison between public and private schools and allows linkages of teacher, school, school district, and principal data. SASS has been administrated four times since 1987–88, most recently in 1999–2000.

SAMPLE SURVEY OF PUBLIC, PRIVATE, CHARTER, AND BIA SCHOOLS

SASS collects data on:

- School districts
- Principals
- Schools
- Teachers
- Library media centers

Purpose

To collect the information necessary for a complete picture of American elementary and secondary education. SASS is designed to provide national estimates for public elementary, secondary, and combined schools and teachers; state estimates of public elementary and secondary schools and teachers; and estimates for private schools and teachers at the national level and by private school affiliation. The focus in 1999–2000 shifted from teacher supply and demand issues to the measurement of teacher and school district capacity. Among the topics examined to measure teacher capacity are teacher qualifications, teacher career paths, and professional development. Among the topics examined to measure school capacity are school organization and decisionmaking, curriculum and instruction, parental involvement, school safety and discipline, and school resources.

Core Components

SASS consists of four core components; these are administered to districts, schools, principals, and teachers. The district questionnaire is sent to a sample of public school districts. The school questionnaires are sent to a sample of public schools and private schools, as well as all charter schools in operation as of 1998–99, and all schools operated by the Bureau of Indian Affairs (BIA) or American Indian/Alaska Native tribes. The principal and teacher questionnaires are sent to a sample of principals and teachers working at the schools which received the school questionnaire. (The Teacher Follow-up Survey is a fifth component, but has its own chapter—see chapter 5.)

School District Survey (formerly titled the Teacher Demand and Shortage Survey—TDS). This survey is mailed to each sampled local education agency (LEA). The respondents are contact people identified by LEA personnel. If no contact person is identified, the questionnaire is addressed to "Superintendent." The School District Questionnaire consists of items about student enrollments, number of teachers, teacher recruitment and hiring practices, teacher dismissals, existence of a teacher union, length of the contract year, teacher salary schedules, school choice, magnet programs, graduation requirements, and professional development for teachers and administrators. The 1999–2000 School District Questionnaire added new items on the percentage of payroll dedicated to school staff benefits, oversight of home-schooled students and charter schools, use of school performance reports, migrant education, and procedures for recruiting and dismissing teachers. Some items that appeared previously have been dropped, such as layoff data and counts of students by grade level (the latter is available through CCD). The School District Questionnaire is mailed only to public school districts. Comparable questions for BIA, charter schools, and private schools appear on those schools' questionnaires.

School Principal Survey (formerly titled the School Administrator Survey). This survey is mailed to principals/heads of schools. The 1999–2000 School Principal Questionnaire appears in four versions: one for principals or heads of public schools, one for heads of private schools, one for heads of charter schools, and one for heads of BIA schools. The four versions contain only minor differences in phrasing to reflect differences in governing bodies and position titles in the schools. The questionnaires collect information about principal/school head demographic characteristics, training, experience, salary, and judgments about the seriousness of school problems. The 1999–2000 School Principal Questionnaire also covers new data on: principals'/school heads' frequency of engaging in various school and school-related activities; perceived degree of influence of principals and other groups (state, local, school, and parents) in setting performance standards for students; barriers (e.g., personnel policies, inadequate documentation, lack of support, stress) to dismissing poor or incompetent teachers; rewards or sanctions for success or failure to meet district or state performance goals; and means for assessing progress on school improvement plans.

School Survey. The SASS School Questionnaire is sent to public schools, private schools, BIA schools, and charter schools. (The Charter School Questionnaire is described below.) School Questionnaires are addressed to "Principal" although the respondent could be any knowledgeable school staff member (e.g., vice principal, head teacher, or school secretary). Items cover grades offered, number of students enrolled, staffing patterns, teaching vacancies, high school graduation rates, programs and services offered, and college application rates. The 1999–2000 version for public, private, and BIA schools incorporates new items on: computers (number, access to the Internet, and whether there is a computer coordinator in the school); availability of certain types of curricular options; how special education students' needs are met; changes in the school year or weekly schedule; the enrollment capacity of schools; and whether schools have programs for disruptive students.

Public Charter School Questionnaire. As a continuation of a national study of charter schools, NCES added a new SASS component on charter schools. All charter schools in operation as of 1998–99 were surveyed in the 1999–2000 SASS. For the first time, there will be comparable data on public, private, BIA, and charter schools. A number of questions that only apply to charter schools are asked, including: when the charter was granted, and by whom; what types of regulations were waived, and their importance; whether the school is new or was converted from a pre-existing school; and whether the school operates within a school district or not. A small number of school library media center items have also been incorporated into the charter school questionnaire, such as whether the school has a library media center, the number of school library media center staff, and the number of students who used the library media center in the past week. Charter schools that operate on their own are asked some of the district items, such as school hiring practices and graduation requirements.

Teacher Survey. This survey is mailed to a sample of teachers from the SASS sample of schools. It is sent out in four versions—to teachers in public schools, private schools, charter schools, and BIA schools. The four versions, however, are virtually identical, except that charter school teachers who worked in the school prior to its becoming a charter school are asked if they supported the conversion. The SASS Teacher Questionnaire collects data from teachers about their education and training, teaching assignment, certification, workload, and perceptions and attitudes about teaching. The 1999–2000 SASS Teacher Questionnaire expands data collection on teacher preparation, induction, organization of classes, and professional development. It also collects data on a new topic: use of computers. The only eligible respondent for each teacher questionnaire is the teacher named on the questionnaire label. As of the 1993–94 SASS, administrators are eligible for both the Teacher Survey and the Principal Survey, if they teach a regularly scheduled class.

Additional Components

In addition to the core data collection described above, SASS featured additional components focusing on library media specialists/librarians and a student records component in 1993–94, and on library media centers in 1993–94 and 1999–2000. One year following each SASS, a Teacher Follow-up Survey (TFS) is mailed to a sample of participants in the SASS Teacher Survey. See chapter 5 for a complete description of TFS.

Library Media Center Survey. This component was added in the 1993–94 SASS. The School Library Media Center Questionnaire asks public, private, and BIA schools about their access to and use of new information technologies. The survey collects data on library collections, media equipment, use of technology, staffing, student services, expenditures, currency of the library collection, and collaboration between the library media specialist and classroom teachers. Schools could respond to the School Library Media Center Questionnaire in the usual paper and pencil mode or by using a web-based survey form on the Internet in 1999–2000. (See chapter 9 for a more complete description of this survey.)

Library Media Specialist/Librarian Survey. This questionnaire was mailed to a subsample of the SASS sample of public, private, and BIA schools in 1993–94. This survey solicited data that could be used to describe school librarians—for example, their educational background, work experience, and demographic characteristics. Because much of the collected information was comparable to that obtained in the Teacher Questionnaires, comparisons between librarians and classroom teachers can be made.

Student Records Component. This questionnaire, along with a roster of sampled students, was mailed to a subsample of the SASS sample of public and private schools in 1993–94. This survey solicited information about a student that could be answered by a school administrator using the student's school record. The information about selected students was not obtained from the students themselves. The survey provided information on the types of services students received, and the types of math and science courses in which they were enrolled. The students can be linked to their schools and teachers.

Periodicity

From 1987–88 to 1993–94, SASS core components were on a 3-year cycle, with the TFS conducted 1 year after SASS. After a 6-year hiatus, SASS was fielded in 1999–2000, with the TFS following in 2000–01. Subsequent SASS administrations are scheduled on a 4-year cycle.

2. USES OF DATA

SASS is the largest, most extensive survey of school districts, schools, principals, teachers, and library media centers in the United States today. It includes data from public, private, and Bureau of Indian Affairs school sectors. Moreover, SASS is the only survey that studies the complete universe of public charter schools. Therefore, SASS provides a multitude of opportunities for analysis and reporting on issues related to elementary and secondary schools.

SASS data have been collected four times over the period between 1987 and 2000. Many questions have been asked of respondents at multiple time points, allowing researchers to examine trends on these topics over time. SASS asks similar questions of respondents across sectors, including public, public charter, Bureau of Indian Affairs, and private schools. The consistency of questions across sectors and the large sample sizes allow for exploration of similarities and differences across sectors.

SASS data are representative at the state level for public school respondents and at the private school affiliation level for private school respondents. Thus, SASS is invaluable for analysts interested in elementary, middle, and secondary schools within or across specific states or private school affiliations. The large SASS sample allows extensive disaggregation of data according to the characteristics of teachers, administrators, school, and school districts. For example, researchers can compare urban and rural settings, and the working conditions of teachers and administrators of differing demographic backgrounds.

SASS collects extensive data on teachers, principals, schools, and school districts. Information on teachers includes their qualifications, early teaching experience, teaching assignments, professional development, and attitudes about the school. School questions include enrollment, staffing, the types of programs and services offered, school leadership, parental involvement, and school climate. At the district level, information is sought on the recruitment and hiring of teachers, professional development programs, student services, and other relevant topics.

SASS data can be very useful for researchers performing their own focused studies on smaller populations of teachers, administrators, schools, or school districts. SASS can supply data at the state, affiliation, or national level that provide valuable contextual information for localized studies; localized studies can provide illustrations of broad findings produced by SASS.

Users of restricted-use SASS data can link school districts and schools to other data sources. For instance, 1999–2000 SASS restricted-use data sets include selected information taken from the NCES Common Core of Data, but researchers can augment the data sets by adding more data from the CCD—either fiscal or nonfiscal data.

3. KEY CONCEPTS

Because of the large number of concepts in SASS surveys, only those pertaining to the level of data collection (LEA, school, teacher, library) are described in this section. For additional terms, the reader is referred to glossaries in SASS reports.

Local Education Agency (LEA). A public school district that is defined as a government agency employing elementary and secondary level teachers and administratively responsible for providing public elementary and/or secondary instruction and educational support services. Districts that do not operate schools but employ teachers are no longer included as of the 1999–2000 SASS. For example, some states have special education cooperatives that employ special education teachers who teach in schools in more than one school district.

Public School. An institution that provides educational services for at least one of grades 1–12 (or comparable ungraded levels), has one or more teachers to give instruction, is located in one or more buildings, receives public funds as primary support, and is operated by an education agency. Schools in juvenile detention centers and schools located on military bases and operated by the Department of Defense are included.

Private School. An institution that is not in the public system and that provides instruction for any of grades 1–12 (or comparable ungraded levels). The instruction must be given in a building that is not used primarily as a private home. Private schools are divided into three categories: (1) Catholic: parochial, diocesan, private order; (2) Other religious: affiliated with a Conservative Christian school association, affiliated with a national denomination, unaffiliated; (3) Nonsectarian: regular, special program emphasis, special education. The three nonsectarian school categories are determined not by governance but by program emphasis. This classification disentangles private schools offering a conventional academic program (regular) from those which either serve special needs children (special education) or provide a program with a special emphasis (e.g., arts, vocational, alternative).

Charter School. A charter school is a public school that, in accordance with an enabling state statute, has been granted a charter exempting it from selected state or local rules and regulations. A charter school may be a newly created school or it may previously have been a public or private school.

BIA School. A school funded by the Bureau of Indian Affairs, U.S. Department of the Interior. These schools may be operated by the BIA, a tribe, a private contractor, or a local education agency (school district).

Library media center (LMC). A library media center is an organized collection of printed, audiovisual, or computer resources that (a) is administered as a unit, (b) is located in a designated place or places, and (c) makes resources and services available to students, teachers, and administrators.

Teacher. A full-time or part-time teacher who teaches any regularly scheduled classes in any of grades K–12.* Includes administrators, librarians, and other professional or support staff who teach regularly scheduled classes on a part-time basis. Itinerant teachers are also included, as well as long-term substitutes who are filling the role of a regular teacher on a long-term basis. An itinerant teacher is one who teaches at more than one school (e.g., a music teacher who teaches three days per week at one school and two days per week at another). Short-term substitute teachers and student teachers are not included.

4. SURVEY DESIGN

Target Population

Local Education Agencies (LEAs) that employ elementary and/or secondary level teachers (e.g., public school districts, state agencies that operate schools for special student populations such as inmates of juvenile correctional facilities, Department of Defense, etc.) and cooperative agencies that provide special services to more than one school district; public, private, BIA, and charter schools with students in any of grades 1–12; principals of those schools, as well as library media centers; and teachers in public, private, BIA, and charter schools who

---

*A teacher teaching only kindergarten students is in scope, provided the school serves students in a grade higher than kindergarten.

BEST COPY AVAILABLE

teach students in grades K–12 in a school with at least a 1st grade.

Sample Design

SASS uses a stratified probability sample design. Details of stratification variables, sample selection, and frame sources are provided below.

Schools are selected first. For the *public school* sample, the first level of stratification is by the five types of school: (a) BIA schools; (b) Native American schools (i.e., schools with 19.5 percent or more Native American students); (c) schools in Delaware, Nevada, and West Virginia (where it is necessary to implement a different sampling methodology to select at least one school from each LEA in the state); (d) charter schools; and (e) all other schools. Schools falling into more than one group are assigned in hierarchical order. In the second level of stratification, Native American schools are stratified by Arizona, California, Minnesota, Montana, New Mexico, North Dakota, Oklahoma, South Dakota, Washington, and all other states (except Alaska, since most Alaskan schools have high Native American enrollment), and schools in Delaware, Nevada, and West Virginia are stratified first by state and then by LEA. Within each second level there were three grade level strata (elementary, secondary, and combined schools).

Within each stratum, all non-BIA and non-Charter schools are systematically selected using a probability proportionate to size algorithm. The measure of size used for the schools on CCD was the square root of the number of teachers in the school as reported on the CCD file. Any school with a measure of size larger than the sampling interval was excluded from the probability sampling operation and included in the sample with certainty.

The Common Core of Data (CCD) Public School Universe serves as the public school sampling frame. (See chapter 2 for a complete description of CCD.) The frame includes regular public schools, Department of Defense-operated military base schools, and special purpose schools such as special education, vocational, and alternative schools. Schools outside the United States and schools that teach only prekindergarten, kindergarten, or postsecondary students are deleted from the file. The following years of CCD were used as the public school frame for the last three rounds of SASS:

- 1997–98 CCD for the 1999–2000 SASS public school sample;
- 1991–92 CCD for the 1993–94 SASS; and
- 1988–89 CCD for the 1990–91 SASS.

In the 1987–88 SASS, the 1986 Quality of Education Data (QED) survey was used as the sampling frame.

For *private schools*, the sample is stratified within each of the two types of frames: (1) a list frame, which is the primary private school frame, and (2) an area frame, which is used to identify schools not included on the list frame and to thereby compensate for the undercoverage of the list frame. For list frame private schools, the schools are stratified by affiliation and school association membership, grade level, and region. All schools in the area frame that are in noncertainty PSUs are included with certainty and those in certainty PSUs are included in the list frame and sampled there. Within each stratum, schools are sampled systematically using a probability proportionate to size algorithm. The measure of size used in 1999–2000 SASS is the square root of the 1997–98 PSS number of teachers in the school. Any school with a measure of size larger than the sampling interval was excluded from the probability sampling process and included in the sample with certainty.

The most recent Private School Survey (PSS), updated with the most recent association lists, serves as the private school sampling frame. For example, the 1997–98 PSS, updated with 26 lists of private schools provided by private school association as well as 51 lists of private schools from the 50 states and the District of Columbia, was used as the private school frame for the 1999–2000 SASS. (See chapter 3 for a complete description of PSS.) The 1991–92 and the 1989–90 PSS were the basis for the private school frame for the 1993–94 and 1990–91 SASS, respectively. The 1986 Quality of Education Data (QED) survey was used as the sampling frame for the 1987–88 SASS.

Since the 1993–94 SASS, all *Bureau of Indian Affairs (BIA) schools* are selected with certainty; in 1990–91, 80 percent of BIA schools were sampled. The Indian School frame for the 1999–2000 SASS consists of a list of schools that the BIA operated or funded during the 1997–98 school year. The list is obtained from the U.S. Department of the Interior. The BIA list is matched against CCD, and the schools on the BIA list which do not match CCD are added to the universe of schools.

A *charter school* frame was added in the 1999–2000 SASS. All charter schools are selected with certainty. The charter school frame consists of a list of charter schools developed for the Institute of Education Sciences (IES). This list includes only charter schools that were open (teaching students) during the 1998–99 school year.

Each sampled school receives a school questionnaire and the principal of each sampled school receives a principal questionnaire.

For the 1999–2000 SASS, as in 1993–94, the ***library media center*** sample was a subsample of the SASS school sample. Each sampled library media center receives a library media center questionnaire.

A sample of ***teachers*** is selected within each sampled school. First, the sampled schools are asked to provide a list of their teachers and selected characteristics. In 1999–2000, teachers were stratified into one of five teacher types in the following hierarchical order: Asian or Pacific Islander; American Indian, Aleut, or Eskimo; Bilingual/English as a Second Language (ESL); New; and Experienced. For new/experienced teachers in public schools, oversampling was not required due to the large number of sample schools with new teachers. Therefore, teachers were allocated to the new and experienced categories proportional to their numbers in the school. However, for private teachers, new teachers were oversampled. Before teachers were allocated to the new/experienced strata, schools were first allocated an overall number of teachers to be selected.

The school-level file that included the number of teachers at the school for the five teacher strata was sorted by school type (public, private, charter), school strata, school order of selection, and school control number. Within each school and teacher stratum, teachers were selected systematically with equal probability. Using the teacher probabilities of selection, take every, and start-withs, sample teachers were selected from each stratum across schools. The within-school probabilities of selection were computed so as to give all teachers within a school stratum the same overall probability of selection (self-weighted). However, since the school sample size of teachers was altered due to the minimum constraint (i.e., at least one teacher/school) or maximum constraint (i.e., no more than either twice the average stratum allocation or 20 teachers/school), the goal of achieving self-weighting for teachers was lost in some schools. Each sampled teacher receives a teacher questionnaire.

Once public schools are selected, the ***districts*** associated with these schools—except in the states of Delaware, Nevada, and West Virginia—are in the sample as well. In Delaware, Nevada, and West Virginia, all districts were defined as school sampling strata, placing all districts in each of these three states in the district sample. (In some SASS administrations a sample of districts not associated with schools is taken, but not in the 1999–2000 SASS.) The district sample is selected using a systematic equal probability algorithm. Each sampled school district receives a school district questionnaire.

The approximate sample sizes for the 1999–2000 SASS are 14,500 schools and administrators, 75,000 teachers, 5,700 school districts, and 13,400 school library media centers.

Data Collection and Processing

The 1999–2000 Schools and Staffing Survey (SASS) was primarily a mailout/mailback survey with computer-assisted telephone interviewing (CATI) and telephone follow up. The School Library Media Center Survey could also be answered through a web-based survey form on the Internet. All survey modes were administered by the U.S. Bureau of the Census.

Reference dates. Data for SASS components are collected during a single school year. Most data items refer to that school year. Questions on enrollment and staffing refer to October 1 of the school year. Questions for teachers about current teaching loads refer to the most recent full week that school was in session, and questions on professional development refer to the past 12 months.

Data collection. The data collection procedures begin with advance mailings to school districts and schools principals explaining the nature and purpose of SASS. The advance mailing to principals includes a request to submit a list of all teachers in their schools. Follow up to the teacher listing form request includes a reminder postcard, a second mailing of the teacher listing form request, and finally telephone calls to all nonrespondents. The teacher sample is selected using these lists.

The school district, principal, and library media center questionnaires are mailed out first, followed by the school questionnaires, and then the teacher questionnaires. Reminder postcards are mailed within 1 to 4 weeks after the initial mailing for each type of questionnaire. A second copy of the questionnaire is mailed to cases that fail to respond to the first mailout within 6 weeks of the reminder postcard.

About 6 weeks after the second mailing for each type of questionnaire, Census Bureau staff members begin telephoning sample units that have not returned questionnaires. Most follow up is done through calls made by Census staff in three centralized locations, using computer-assisted telephone interviewing (CATI) to collect the questionnaire data.

BEST COPY AVAILABLE

Finally, nonrespondent school districts, private schools, BIA schools, charter schools, and public and private school teachers are called or visited by field representatives (FRs). These FRs complete paper copies of the questionnaires as they collect the data. In some cases where the respondent is unwilling to participate in an interview, the FR attempts to persuade him/her to return a mailed questionnaire. (Due to budgetary constraints, FRs collected data from a subsample of public and private school teacher nonrespondents in 1999–2000.)

Processing. As of the 1999–2000 SASS, imaging technology was used instead of data keying. After data entry, the files of scanned data from paper questionnaires are merged with those from the computer-assisted telephone interviews (CATI). The next step is to make a preliminary determination of each case's interview status (ISR); that is, whether it is an interview, a noninterview, or out of scope. Then interview records on the data files are processed through a computer pre-edit program designed to identify inconsistencies and invalid entries. Census staff reviews the problem cases and make corrections whenever possible.

After pre-edit corrections are made, all records (i.e., from all survey components) classified as interviews at this point are subject to a set of computer edits: a range check, a consistency edit, and a blanking edit. After the completion of these edits, the records are put through another edit to make a final determination of whether the case is eligible for the survey, and, if so, whether sufficient data have been collected for the case to be classified as an interview. A final interview status recode (ISR) value is assigned to each case as a result of the edit.

Estimation Methods

Sample units are weighted to produce national and state estimates for public elementary and secondary school surveys (i.e., schools, teachers, administrators, school districts, and school library media centers); and national estimates for BIA, charter school, and public "combined" school surveys (i.e., schools, teachers, administrators, and school library media centers). The private sector is weighted to produce national and affiliation group estimates. These estimates are produced through the weighting and imputation procedures discussed below.

Weighting. Estimates from SASS sample data are produced by using weights. The weighting process for each component of SASS includes adjustment for nonresponse using respondents' data, and adjustment of the sample totals to the frame totals to reduce sampling variability. The exact formula representing the construction of the weight for each component of SASS is provided in each administration's sample design report (e.g., *1993–94 Schools and Staffing Survey: Sample Design and Estimation*, NCES 96–089). The construction of weights is also discussed in the *Quality Profiles* (NCES 2000–308 and NCES 94–340). Since data for SASS were collected at the same time as for PSS in 1993–94 and 1999–2000, in both those years the number of private schools reported in SASS was made to match the number of private schools reported in PSS.

Imputation. In all administrations of SASS, all item missing values are imputed for records classified as interviews. SASS uses a two-stage imputation procedure. The first stage imputation process uses a logical or deductive type of imputation method, such as:

(1) Using data from other items on the same questionnaire;

(2) Extracting data from a related SASS component (different questionnaire); and

(3) Extracting information about the sample case from the Private School Survey or the Common Core of Data, the sampling frames for private and public schools.

In addition, some inconsistencies between items are corrected by ratio adjustment during the first stage imputation.

The second stage imputation process is applied to all items with missing values that were not imputed in the first stage. This imputation uses a hot-deck imputation method, extracting data from a respondent (donor) with similar characteristics to the nonrespondent. If there is still no observed value after collapsing to a certain point, the missing values are imputed by clerical imputation.

Recent Changes

During the 6-year hiatus between the 1993–94 SASS and the 1999–2000 SASS, a redesign effort was undertaken. NCES involved various programs in the Department of Education and the wider education research and policy community in the planning process for the SASS redesign.

Design changes from 1993–94 to 1999–2000:

- For the private sector, the sample was reallocated to publish estimates for one additional association, making a total of 20 associations.
- A list of Department of Defense (DOD) schools was obtained and included on the sampling frame giving SASS complete coverage of domestic DOD schools.

- The Department of Education, Institute of Education Sciences (IES), provided a list of public charter schools, giving SASS coverage of charter schools open in the 1998–1999 school year. Questionnaires were prepared to include some items specific to charter schools.

- The variance methodology was altered: in earlier SASS administrations, it was assumed that there was no variance associated with certainty schools, and that all error from certainty schools reflected bias. In 1999–2000, it was decided to assume that nonresponse from certainty schools followed a random process and so certainty schools could have variance due to this random process.

- Additional size classes were introduced into all weighting procedures and were customized by state and private school association.

- The control of the overlap with the previous SASS was dropped and replaced with a procedure designed to minimize the overlap between SASS and National Assessment of Educational Progress (NAEP) sample schools.

- The bootstrap variance system was refined to produce more stable variance estimates.

- The LMC sample size was first expanded to include all SASS schools and then, for cost and burden reasons, reduced to exclude charter schools. The charter school questionnaire included a small selection of questions from the LMC questionnaire.

Content changes from 1993–94 to 1999–2000. For the 1999–2000 school year, these components were dropped from SASS:

- The Library Media Center Specialist/Librarian component of the 1993–94 SASS was dropped.

- The student records component of the 1993–94 SASS was dropped.

Changes were also made to existing SASS components, based on two extensive field tests.

- *Additions to School Questionnaire*: number of computers, access to the Internet, whether there is a computer coordinator in this school, availability of certain types of curricular options, how special education students' needs are met, changes in the school year or weekly schedule, the enrollment capacity of schools, and whether schools have programs for disruptive students. A charter school questionnaire was added to this series; it included elements of the District and Library Media Center Questionnaire since those two components did not add a separate charter school questionnaire.

- *Deletions to School Questionnaire*: layoff data and counts of students by grade level.

- *Additions to Principal Questionnaire*: principals'/school heads' frequency of engaging in various school and school-related activities, perceived degree of influence of principals and other groups (state, local, school, and parents) in setting performance standards for students, barriers (e.g., personnel policies, inadequate documentation, lack of support, stress) to dismissing poor or incompetent teachers, rewards or sanctions for success or failure to meet district or state performance goals, and means for assessing progress on school improvement plan. A charter school questionnaire was added to this series.

- *Deletions to Principal Questionnaire*: degrees earned—other than highest (including their dates, in what field they were earned, and at which college or university a bachelor's degree was earned), the location and grade levels of the previous school at which respondent was principal, breaks in service, year when eligible to retire, and benefits received in addition to salary.

- *Additions to Teacher Questionnaire*: training, teacher induction, teacher professional development, curriculum development, computer usage and decisionmaking practices. A charter school questionnaire was added to this series.

- *Additions to School District Questionnaire*: percentage of payroll dedicated to school staff benefits, oversight of home-schooled students and charter schools, use of school performance reports, migrant education, and procedures for recruiting and dismissing teachers.

Internet reporting option. In addition to the paper SASS forms, an Internet reporting option was developed for the public and private Library Media Center Questionnaire.

Questionnaire printing. The 1999–2000 SASS was the first administration of SASS to use customized printing of questionnaires. For SASS, it was used to:

- Print respondent's identification information on any page.

- Provide information to specific respondents to avoid definitional problems.

- Split-panel wording for an LMC test.

- Personalize letters to respondents.

Future Plans

SASS administrations are now scheduled on a 4-year cycle. The next administration will be in 2003–2004.

BEST COPY AVAILABLE

5. Data Quality and Comparability

Sampling Error

The estimators of sampling variances for SASS statistics take the SASS complex sample design into account. For an overview of the calculation of sampling errors, see the *SASS Quality Profiles* (NCES 2000–308 and NCES 94–340).

Direct variance estimators. The *balanced half-sample replication (BHR)* method, also called *balanced repeated replication (BRR)* method, was used to estimate the sampling errors associated with estimates from the 1987–88 and 1990–91 SASS. Given the replicate weights, the statistic of interest (such as the number of 12th grade teachers from the School Survey) can be estimated from the full sample and from each replicate. The mean square error of the replicate estimates around the full sample estimate provides an estimate of the variance of the statistic.

A bootstrap variance estimator was used for the 1993–94 and the 1999–2000 SASS. The bootstrap variance reflects the increase in precision due to large sampling rates because the bootstrap is done systematically without replacement, as was the original sampling. Bootstrap samples can be selected from the bootstrap frame, replicate weights computed, and variances estimated with standard BHR software. The bootstrap replicate basic weights (inverse of the probability of selection) were subsequently reweighted. For more information about the bootstrap variance methodology and how it applies to SASS see: "A Bootstrap Variance Estimator for Systematic PPS Sampling" in NCES Working Paper 2000–04, *Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings* (this paper describes the methodology used in 1999–2000 SASS), "A Bootstrap Variance Estimator for the Schools and Staffing Survey" and "Balanced Half-sample Replication with Aggregation Units" in NCES Working Paper 94–01, *Schools and Staffing Survey (SASS), Papers Presented at the Meetings of the American Statistical Association*; "Comparing Three Bootstrap Methods for Survey Data" by Randy Sitter, in the Technical Report Series of the Laboratory for Research in Statistics and Probability, published by Carleton University in 1990; "Properties of the Schools and Staffing Survey Bootstrap Variance Estimator" in NCES Working Paper 96–02, *Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association*; and "The Jackknife, the Bootstrap and other Resampling Plans," an article by Bradley Efron in *Society for Industrial and Applied Mathematics*, (SIAM) No. 38, 1982.

The replicate weights for all three rounds of SASS are used to compute the variance of a statistic, Y, as stated below.

$$Variance(Y) = \frac{1}{n}\sum_{r}(Y_r - Y)^2$$

where: Y_r = the estimate of Y using the rth set of replicate weights, and

n = the number of replicates (n=88 for 1999–2000 SASS).

SASS variances can be calculated using the 88 replicates of the full sample that are available on the data files with software such as WesVarPC. For examples of other software that support BRR, see K.M. Wolter's *Introduction to Variance Estimation* (New York: Springer-Verlag, 1985).

Average design effects. Design effects (*Deff*) measure the impact of the complex sample design on the accuracy of a sample estimate, in comparison to the alternative simple random sample design. For the 1990–91 SASS, an average design effect was derived for groups of statistics, and within each group, for a set of subpopulations. Standard errors of 1990–91 and 1993–94 SASS statistics of various groups for various subpopulations can then be calculated approximately from the standard errors based on the simple random sample (using SAS or SPSS) in conjunction with the average design effects provided. For example, average design effects for selected variables in the Public School Survey are 1.60 (public sector) and 1.36 (private sector); in the Principal Survey, 4.40 (public sector) and 4.02 (private sector), and in the Teacher Survey, 3.75 (public sector) and 2.52 (private sector). Examples illustrating the use of SASS average design effect tables are provided in *Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Survey (SASS), Volume I, User's Manual* (NCES 95–342–I).

Generalized variance functions (GVF). GVF tables were developed for use in the calculation of standard errors of totals, averages, and proportions of interest in the 1990–91 SASS components. The 1990–91 GVFs can be used for the 1993–94 SASS because no major design changes were adopted between 1990–91 and 1993–94. Examples illustrating the use of the GVF tables are provided in *Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Survey (SASS), Volume I, User's Manual* (NCES 95–342–I). Note that the GVF approach, unlike the design effect approach described above, involves no need to calculate the simple random sample variance estimates.

Nonsampling Error

Coverage error. SASS surveys are subject to any coverage error present in CCD and PSS, the NCES data files that serve as their principal sampling frames. The report *Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary Education Agency Universe Survey* (NCES 97-505) found that overall coverage in the Agency Universe Survey was 96.2 percent (in a comparison to state education directories). "Regular" agencies—those traditionally responsible for providing public education—had almost total coverage in the 1994–95 survey. Most coverage discrepancies were attributed to nontraditional agencies that provide special education, vocational education, and other services. However, there is potential for undercoverage bias associated with the absence of schools built between the construction of the sampling frame and time of the SASS survey administration. Further research on coverage can be found in "Evaluating the Coverage of the U.S. National Center for Education Statistics' Public Elementary/Secondary School Frame" (Hamann 2000) and "Evaluating the Coverage of the U.S. National Center for Education Statistics' Public and Private School Frames Using Data from the National Assessment of Educational Progress" (Lee, Burke, and Rust 2000).

A capture-recapture methodology was used to estimate the number of private schools in the United States and to estimate the coverage of private schools in the 1999–2000 PSS; the study found that the PSS school coverage rate is equal to 97 percent. (See chapter 2 for a description of CCD and chapter 3 for a description of PSS.)

Nonresponse error.

Unit nonresponse. The weighted unit response rates for public schools have been higher than the weighted unit response rates for private schools in the first three rounds of SASS (rates for 1999–2000 are not available at this time). See table 2. For more information on the analysis of nonresponse rates, refer to *An Analysis of Response Rates in the 1993–94 Schools and Staffing Survey* (NCES 98-243) and *An Exploratory Analysis of Response Rates in the 1990–91 Schools and Staffing Survey (SASS)* (NCES 96-338).

Item nonresponse. The percentage of items with response rates of 90 percent or more was generally high across the first three rounds of SASS (rates for 1999–2000 are not available at this time): for example, in 1993–94, for public schools, 91 percent of the School District Surveys had item response rates of 90 percent or more, 92 percent of Principal Surveys, 83 percent of School Surveys, and 91 percent of Teacher Surveys. Item response rates gradually increased between 1987–88 and 1993–94. They ranged from 11 to 100 percent in the 1987–88 SASS, 25 to 100 percent in the 1990–91 SASS, and 50 to 100 percent in the 1993–94 SASS. (See the *SASS Data File User's Manuals*, NCES 96–142 and NCES 93–144–I.)

Measurement error. Results reported in *An Analysis of Response Rates in the 1993–94 Schools and Staffing Survey* (NCES 98–243) support the contention that, without follow up to mail surveys, nonresponse error would be much greater than it is and that the validity and reliability of the data would be considerably reduced. However, because of the substantial amount of telephone follow up, there is concern about possible bias due to differences in the mode of survey collection. Other possible sources of measurement error include long, complex instructions that respondents either do not read or do not understand, navigation problems related to the format of the questionnaires, and definitional and classification problems. See also *Measurement Error Studies at the National Center for Education Statistics* (NCES 97-464).

Table 2. Summary of overall weighted unit response rates for selected SASS questionnaires

Questionnaire	1987–88	1990–91	1993–94
School District Survey	90.8	93.5	93.9
Public Principal Survey	94.4	96.7	96.6
Public School Survey	91.9	95.3	92.3
Public Teacher Survey*	82.9	85.9	83.8
Private Principal Survey	79.3	90.1	87.6
Private School Survey	78.6	83.9	83.2
Private Teacher Survey*	69.6	75.5	72.9
BIA Principal Survey	†	†	98.7
BIA School Survey	†	†	99.3
BIA Teacher Survey	†	†	86.5

†Not applicable

*The overall teacher response rates are the percentage of teachers responding in schools that provided teacher lists for sampling. The response rates to the Public Teacher Survey itself ranged from 86.4 (in 1987–88) to 90.3 percent (in 1990–91) and to the Private Teacher Survey from 79.1 (in 1987–88) to 83.6 percent (in 1990–91).

SOURCE: Choy, Medrich, and Henke, *Schools and Staffing in the United States: A Statistical Profile, 1987–88* (NCES 92-120). Gruber, *1990–91 Schools and Staffing Survey: Data File User's Manual* (NCES 93–144-I). Gruber, Rohr, and Fondelier, *1993-94 Schools and Staffing Survey: Data File User's Manual* (NCES 96–142). Jabine, *Quality Profile for SASS: Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS)* NCES 94–340.

BEST COPY AVAILABLE

Several NCES working papers also address measurement error. Reports that study the 1993–94 SASS include: *Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey* (NCES Working Paper 96–05); *Further Cognitive Research on the Schools and Staffing Survey (SASS)* (NCES Working Paper 97–23); *Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993–94 School Year* (NCES Working Paper 97–10), and Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report (NCES Working Paper 98–02). Reports that study the 1991–92 SASS include: *1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report* (NCES Working Paper 94–03) and *The Results of the 1991–92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation* (NCES Working Paper 98–02).

6. CONTACT INFORMATION

For content information on SASS, contact:

Kathryn Chandler
Phone: (202) 502–7486
E-mail: kathryn.chandler@ed.gov
SASS e-mail: sassdata@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

1987–88 Schools and Staffing Survey - Public School Administrator Questionnaire Data, NCES 91–137, by P. Broene. Washington, DC: 1991.

1987–88 Schools and Staffing Survey - Public School Data, NCES 91–136, by P. Broene. Washington, DC: 1991.

1990–91 Schools and Staffing Survey: Data File User's Manual Volume I: Survey Documentation, NCES 93–144–I, by K. Gruber. Washington, DC: 1994.

1993–94 Schools and Staffing Survey: Data File User's Manual, Volume I: Survey Documentation, NCES 96–142, by K. Gruber, C. Rohr, and S. Fondelier. Washington, DC: 1996.

What Users Day About Schools and Staffing Survey Publications, NCES Working Paper 1999–10, by U. Rouk, L. Weiner, and D. Riley. Washington, DC: 1999.

Uses of Data

An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey, NCES Working Paper 98–18, by R.M. Ingersoll. Washington, DC: 1998.

A Research Agenda for the 1999–2000 Schools and Staffing Survey, NCES Working Paper 2000–10, by D.J. McGrath and M.T. Luekens. Washington, DC: 2000.

The Schools and Staffing Survey: Recommendations for the Future, NCES 97–596, by J.E. Mullens and D. Kasprzyk. Washington, DC: 1997.

Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results, NCES Working Paper 1999–02, by S.D. Wiley and K.A. Reynolds. Washington, DC: 1999.

Survey Design

1987–88 Schools and Staffing Survey: Sample Design and Estimation, NCES 91–127, by S. Kaufman. Washington, DC: 1991.

1991 Schools and Staffing Survey: Sample Design and Estimation, NCES 93–449, by S. Kaufman. Washington, DC: 1993.

1993–94 Schools and Staffing Survey: Sample Design and Estimation, NCES 96–089, by R. Abramson, C. Cole, S. Fondelier, B. Jackson, R. Parmer, and S. Kaufman. Washington, DC: 1996.

An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey, NCES Working Paper 95–18, by R.M. Ingersoll. Washington, DC: 1995.

Collection of Public School Expenditure Data: Development of a Questionnaire, NCES Working Paper 98–01, by J.B. Isaacs, C.M. Best, A.D. Cullen, M.S. Garet, and J.D. Sherman. Washington, DC: 1998.

Collection of Resource and Expenditure Data on the Schools and Staffing Survey, NCES Working Paper 1999–07, by J.B. Isaacs, M.S. Garet, J.D. Sherman, A. Cullen, and R. Phelps. Washington, DC: 1999.

A Feasibility Study of Longitudinal Design for Schools and Staffing Survey, NCES Working Paper 98–16, by D. Baker, R. Levine, M. Han, and M. Garet. Washington, DC: 1998.

Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey, NCES Working Paper 97–42, by R.E. Levine, J.G. Chambers, I.E. Duenas, and C.S. Hikido. Washington, DC: 1997.

National Assessment of Teacher Quality, NCES Working Paper 96–24, by R.M. Ingersoll. Washington, DC: 1996.

The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper, NCES Working Paper 98–08, by M. Rollefson. Washington, DC: 1998.

Data Quality and Comparability

An Analysis of Response Rates in the 1993–94 Schools and Staffing Survey, NCES 98–243, by D. Monaco, S. Salvucci, F. Zhang, and M. Hu. Washington, DC: 1997.

Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey, NCES Working Paper 96–05, by C.R. Jenkins and D. Von Thurn. Washington, DC: 1996.

Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Surveys (SASS) Volume I—User's Manual, NCES 95–342–I, by S. Salvucci, S. Weng, and S. Kaufman. Washington, DC: 1995.

Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Survey (SASS): Volume II, Technical Report, NCES 95–342–II, by S. Salvucci, A. Holt, R. Moonesinghe, and S. Kaufman. Washington, DC: 1995.

"Evaluating the Coverage of the U.S. National Center for Education Statistics' Public and Private School Frames Using Data from the National Assessment of Educational Progress," *The Second International Conference on Establishment Surveys: Survey Methods for Businesses, Farms, and Institutions* (pp. 89–98), by H. Lee, J. Burke, and K. Rust. Arlington, VA: American Statistical Association, 2000.

"Evaluating the Coverage of the U.S. National Center for Education Statistics' Public Elementary/Secondary School Frame," *The Second International Conference on Establishment Surveys: Survey Methods for Businesses, Farms, and Institutions* (pp. 79–88), by T.A. Hamann. Arlington, VA: American Statistical Association, 2000.

An Exploratory Analysis of Response Rates in the 1990–91 Schools and Staffing Survey (SASS), NCES 96–338, by F. Scheuren, D. Monaco, F. Zhang, G. Ikosi, and M. Chang. Washington, DC: 1996.

Further Cognitive Research on the Schools and Staffing Survey (SASS), NCES Working Paper 97–23, by A. Zukerberg and M. Lee. Washington, DC: 1997.

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, and M. Saba. Washington, DC: 1997.

Quality Profile for SASS: Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS), NCES 94–340, by T. Jabine. Washington, DC: 1994.

Quality Profile for SASS Rounds 1–3: 1987–1995, Aspects of the Quality of Data in the Schools and Staffing Surveys (SASS), NCES 2000–308, by G. Kalton, M. Winglee, S. Krawchuck, and D. Levine. Washington, DC: 2000.

Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey, NCES Working Paper 97–10, by C.R. Jenkins. Washington, DC: 1997.

Chapter 5: SASS Teacher Follow-up Survey (TFS)

1. OVERVIEW

SAMPLE FOLLOW-UP SURVEY OF PUBLIC, PRIVATE, CHARTER, AND BIA SCHOOL TEACHERS

SASS collects data on:
- Stayers
- Movers
- Leavers

The SASS Teacher Follow-up Survey (TFS) is a follow-up survey of elementary and secondary school teachers who participated in the Schools and Staffing Survey (SASS, see chapter 4). TFS is conducted for NCES by the U.S. Bureau of the Census in the school year following the SASS data collection. TFS consists of all sampled teachers who left teaching within the year after the SASS was administered and a subsample of those who continued teaching, including those who remained in the same school as in the previous year and those who changed schools.

Purpose

To provide estimates of teacher attrition, retention, and mobility in public and private schools and to project demand for teachers; to provide national data on the characteristics of teachers who leave teaching, their reasons for leaving, and their current occupational status; and to provide information on the career paths of persons who remain in teaching. TFS is designed to support estimates of public elementary, secondary, and combined school teachers and private school teachers at the national level.

Components

TFS is comprised of two questionnaires: one for those who leave the teaching profession (former teachers), and one for those who remain in the teaching profession. These questionnaires ask teachers about their current status, occupational changes and plans, reasons for staying in (or leaving) teaching, and attitudes about the teaching profession. Eligible survey respondents are teachers in public, public charter (as of 2000–2001), private, and Bureau of Indian Affairs (BIA) elementary and secondary schools in the 50 states and the District of Columbia.

Teacher Followup Survey Questionnaire for Former Teachers. This questionnaire collects information on former teachers to ascertain information on current occupation; primary activity; plans to remain in current position; plans for further education, plans for returning to teaching; reasons for leaving teaching; possible areas of satisfaction or dissatisfaction with teaching; salary; marital status; number of children; and other information that may be related to attrition; and reasons for retirement.

Teacher Followup Survey Questionnaire for Continuing Teachers. This questionnaire collects information on continuing teachers to ascertain occupational status (full-time, part-time); primary teaching assignment by field; teaching certificate; level of students taught; areas of satisfaction or dissatisfaction; new degrees earned or pursued; expected duration in teaching; marital status; number of children; academic year base salary; time spent performing school related tasks; use of technology for teaching and learning; effectiveness of school administration; and reasons for leaving previous school.

Periodicity

The first administration of TFS was in the 1988–89 school year with a sample from the 1987–88 SASS of about 2,500 teachers who had left teaching and 5,000 who were still in teaching. The size of the sample is approximately the same for every cycle of TFS. There have been three more administrations of TFS, 1991–92 and 1994–95, and 2000–2001. Each collection of TFS is a follow up to the SASS sample of the previous year.

2. USES OF DATA

Data from TFS are used for a variety of purposes by Congress, state education departments, federal agencies, private school associations, teacher associations, and educational organizations. TFS can be used to address issues related to teacher turnover. Leavers, movers, and stayers can be profiled and compared in terms of teaching qualifications, working conditions, attitudes toward teaching, job satisfaction, salaries, benefits, and other incentives and disincentives for remaining in or leaving the teaching profession. TFS also provides a measure of national teacher attrition in the various fields and updates information on the education, other training, and career paths of teachers. In addition, sampled teachers can be linked to SASS data to determine relationships between local district and school policies/practices, teacher characteristics, and teacher attrition and retention.

3. KEY CONCEPTS

For additional terms, see the glossaries in TFS reports, in particular *Characteristics of Stayers, Movers, and Leavers: Results from the Teacher Followup Survey: 1994–95* (NCES 97–450).

Leavers. Teachers who left the teaching profession in the year after the last SASS administration.

Movers. Teachers who were still teaching in the year after the last SASS administration but had moved to a different school.

Stayers. Teachers who were teaching in the same school in the year after the last SASS administration as in the year of the SASS administration.

Itinerant Teacher. An individual who teaches at more than one school; for example, a music teacher who teaches three days per week at one school and two days per week at another.

4. SURVEY DESIGN

Target Population

The universe of elementary and secondary school teachers who teach in public, private, public charter (as of 1999–2000), and BIA schools in the 50 states and the District of Columbia in schools that had any of grades 1–12 during the school year of the last SASS administration. This population is divided into two components—those who left teaching after that school year (former teachers) and those who continued teaching (current teachers).

Sample Design

TFS surveys a sample of teachers who were interviewed in the previous SASS Teacher Survey. The TFS sample is a stratified sample allocated to allow comparisons of stayers, movers, and leavers by sector, experience, and teaching level. The sample is stratified in the following order: (1) Sector (public, private, and, as of the 2000–2001 TFS, charter); (2) Teacher status (leavers, stayers, movers, unknown); (3) Experience (new/experienced); and (4) Teaching level (elementary, secondary).

Within each public TFS stratum, teachers who respond to the previous SASS Teacher Survey are sorted by subject (i.e., the subject that the teacher teaches the most classes in), Census region, urbanicity, school enrollment, and SASS teacher control number. Within each private TFS stratum, responding teachers are sorted by subject, association membership (list frame), affiliation (area frame), urbanicity, school enrollment, and SASS teacher control number.

After they are sorted, teachers are selected within each stratum using a probability proportional to size (pps) sampling procedure. The measure of size is the teacher weight for the previous SASS. (Note that the SASS teacher weight used in 1993–94 did not include a teacher adjustment factor—a ratio adjustment to the school questionnaire report of teacher head counts—since the TFS sampling needed to be completed before the SASS teacher weight was finalized. See *1993–94 Schools and Staffing Survey: Sample Design and Estimation*, NCES 96–089.)

The 1994–95 TFS surveyed approximately 7,200 teachers who had been interviewed in the 1993–94 SASS Teacher Survey. (See chapter 4 for information on the SASS sample design.) A total of 5,025 public school teachers, 2,098 private school teachers, and 50 BIA school teachers were selected, of whom 4,528, 1,751, and 44, respectively, were interviewed. The target sample sizes for the 2000–2001 TFS include 4,900 stayers and 3,400 leavers.

Data Collection and Processing

The TFS is conducted using mailed questionnaires with telephone follow up. The U.S. Bureau of the Census is the collection agent.

Reference dates. Most data items refer to teacher status at the time of questionnaire completion. Some items refer to the past school year, past semester, past 12 months, or the next school year.

Data collection. In September of the year of survey administration, the Census Bureau mails teacher status forms to schools that provided lists of teachers for the previous SASS. On this form, the school principal (or other knowledgeable staff member) is asked to report the current occupational status of each teacher who was sampled in the previous SASS by indicating whether he/she is still at the school in a teaching or nonteaching capacity, or left the school to teach elsewhere or for a nonteaching occupation. If school staff indicate a sample teacher has moved, the Census Bureau tries to obtain the correct home address from the U.S. Postal Service.

The following January, the TFS questionnaires are mailed to selected teachers and former teachers. The Questionnaire for Former Teachers is sent to sample persons reported by school administrators as having left the teaching profession. The Questionnaire for Current Teachers is sent to sample persons who are reported as still teaching at the elementary or secondary level. The questionnaires are mailed to home addresses when available. Otherwise, they are mailed to the sample teacher's school as listed in the previous SASS administration.

In February, the Census Bureau mails a second questionnaire to each sample person who did not return the first questionnaire. Also, for those who returned the first form and indicated that it does not apply to them (because their status was incorrectly reported by their school in the last SASS administration), the appropriate questionnaire is mailed to them at this time.

In late March, Census interviewers begin calling sample persons who did not return a mail questionnaire. In addition to these nonresponse follow-up cases, some "nonmailable" cases (cases with incomplete addresses) are assigned for telephone follow up. If the interviewers are unable to contact a sample teacher through a contact person or through directory assistance, they call the sample person's school to obtain information about the person's current address or employer. Interviewers use the Telephone Questionnaire for the Teacher Followup Survey to collect the data. This allows the data for current and former teachers to be recorded on the same form. Telephone follow up of nonrespondents is completed by the end of the school year.

Editing. Questionnaires undergo several stages of editing. Upon receipt, clerks assign codes to each questionnaire to indicate its status (e.g., complete interview, refusal, deceased) and then perform a general clerical edit that includes reviewing all entries for legibility and making corrections. For the Questionnaire for Former Teachers, clerks assign industry and occupation codes to the respondent's current job. For the Questionnaire for Continuing Teachers, respondents teaching in a different state are assigned a new state FIPS code.

Once the data are keyed, the next step is to make a preliminary determination of each case's interview status—that is, whether it is an interview, a noninterview, or out-of-scope for the survey. The data file is then divided into two files: (1) former teachers (leavers) and (2) current teachers (stayers and movers). Records classified as interviews in the preliminary interview status check are then submitted to a series of computer edits: range checks, consistency edits, and blanking edits. Next, the records undergo a final edit to determine whether the case is eligible for inclusion in the survey and, if so, whether sufficient data have been collected for the case to be classified as an interview. A final interview status recode (ISR) value is then assigned to each case.

Estimation Methods

Estimates from TFS sample data are produced using weighting and imputation procedures.

Weighting. The TFS weighting process includes adjustment for nonresponse using respondents' data and adjustment of the sample totals to the frame totals to reduce sampling variability. The exact formula for TFS weight construction is provided in *1993–94 Schools and Staffing Survey: Sample Design and Estimation* (NCES 96–089).

Imputation. In all administrations of TFS, all item missing values are imputed for records classified as interviews. Values are imputed by using data from (1) other items on the questionnaire or the previous SASS Teacher Survey record for the same respondent, or (2) data from the record for a respondent with similar characteristics (commonly known as the nearest neighbor "hot-deck" method for imputing for item nonresponse).

Although most imputation is carried out through computer processing, there are some cases where entries are clerically imputed for a few items. In these cases, the data record, the SASS teacher file record, and in some cases, the questionnaire are reviewed, and an entry consistent with the information from those sources is imputed. This procedure is used when (1) there is not a suitable record to use as a donor, (2) the computer method produces an entry that is outside the acceptable range for the item, or (3) there are very few cases where an item is unanswered (usually less than 10).

Recent Changes

Changes between the 1994–95 and 2000–2001 TFS include new items added to measure the impact of retirement policies on teacher supply and the addition of items on general instructional practices across elementary, secondary, and combined schools, particularly as they pertain to the use of computers and other technology in schools. The teacher time use section was also expanded to measure specific demands on teacher time. In some cases, the number of response categories were collapsed for the 2000–01 TFS in response to results of focus group analysis, and several items were slightly altered from the 1994–95 TFS to make them more consistent with the comparable items from the 1999–2000 SASS Teacher Questionnaire.

Future Plans

After a 6-year hiatus, SASS was fielded in 1999–2000, and TFS in 2000–2001. Subsequent administrations are scheduled on a 4-year cycle.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Since the TFS sample is a proper subsample of the SASS teacher sample, the SASS teacher replicates are used for the TFS sample. See the discussion of sampling error and variance estimation in chapter 4 on SASS. In the case of TFS, the TFS basic weight for each TFS teacher is multiplied by each of the SASS replicate weights (n=48 for the 1993–94 SASS; n=88 for the 1999–2000 SASS) divided by the SASS teacher full-sample intermediate weight for that teacher. To calculate the replicate weights which should be used for variance calculations, these TFS replicate basic weights are processed through the remainder of the TFS weighting system.

Nonsampling Error

Coverage error. A potential bias may be introduced into TFS because the TFS frame only includes teachers who responded to SASS.

Nonresponse error.

Unit nonresponse. The total weighted response rate in the 1994–95 TFS was 91.6 percent. Rates were similar for current and former teachers: 91.8 percent for current teachers and 88.8 for former teachers. There was greater variation by school type, with private schools generally having lower response rates than public and BIA schools (87.2 percent versus 92.3 and 99.5 percent, respectively).

Cumulative overall response rates for TFS surveys are the product of the SASS Teacher List response rate, the SASS Teacher Survey response rate, and the TFS Teacher response rate. (See table below.)

Table 3. Weighted response rates for 1993–94 SASS Teacher List, 1993–94 SASS Teacher Survey, 1994–95 TFS, and the cumulative overall response rates

	Sector			
	Public		Private	
SASS Teacher List response rate[1]	95.0		91.0	
SASS Teacher Survey response rate[2]	[3]88.2		[4]80.2	
	Current teachers[5]	Former teachers	Current teachers	Former teachers
Teacher Follow-up Survey response rate[6]	92.5	89.2	87.2	87.6
Cumulative overall response rate	77.5	74.7	63.6	63.9

[1] Weighted percent of schools providing teacher lists for the 1993–94 SASS Teacher Survey.
[2] Weighted percent of eligible sample teachers responding to the 1993–94 SASS Teacher Survey.
[3] This rate does not include the 5 percent of the public schools that did not provide teacher lists.
[4] This rate does not include the 9 percent of the private schools that did not provide teacher lists.
[5] Includes stayers and movers.
[6] Weighted percent of eligible sample teachers responding to the 1994–95 Teacher Follow-up Survey.

SOURCE: Whitener, Gruber, Rohr, and Fondelier, *1994–95 Teacher Followup Survey Data File User's Manual Restricted-Use Version* (NCES Working Paper 1999–14).

BEST COPY AVAILABLE

Item nonresponse. Several items in the 1994–95 TFS had a response rate of less than 80 percent. In the Teacher Followup Survey Questionnaire for Former Teachers, the item asking years to retirement had a response rate below 80 percent. In the Teacher Followup Survey Questionnaire for Current Teachers, items with response rates below 80 percent included one item on type of certificate held in field, three items referring to before-tax earning from teaching and other employment during the summer of 1994, two items on jobs outside the school system during the current school year, and an item on the number of dependents other than spouse and children.

Measurement error. Reinterviews were conducted for the purpose of measuring response variance in the 1994–95 TFS. The reinterview was conducted through two reinterview questionnaires—one for mail cases and another for telephone cases. Each questionnaire contained a subset of questions from the original questionnaire. Seventy-eight percent of the questions evaluated displayed high response variance; only 5 percent displayed low response variance (all but one of the 54 questions on teaching methods had moderate or high response variance). This reinterview study again confirmed that "mark all that apply" questions tend to be problematic. See *Response Variance in the 1994–95 Teacher Follow-up Survey* (NCES Working Paper 98–13). A similar reinterview study is planned for the 2000–01 TFS.

Data Comparability

Caution must be used in the interpretation of change estimates between the TFS surveys prior to 1994–95 and those of 1994–95 and later because of wording changes in the TFS surveys.

6. CONTACT INFORMATION

For content information on TFS, contact:

Kathryn Chandler
Phone: (202) 502–7486
E-mail: kathryn.chandler@ed.gov

Kerry Gruber
Phone: (202) 502–7349
E-mail: kerry.gruber@ed.gov

SASS e-mail: sassdata@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

1991–92 SASS Teacher Follow-up Survey Data File User's Manual Public-Use Version, NCES 94–331, by S.D. Whitener. Washington, DC: 1994.

1991–92 SASS Teacher Follow-up Survey Data File User's Manual Restricted-Use Version, NCES 94–478, by S.D. Whitener. Washington, DC: 1994.

1994–95 Teacher Follow-up Survey, Data File User's Manual Public-Use Version, NCES 98–232, by S.D. Whitener, K.J. Gruber, C. Rohr, and S. Fondelier. Washington, DC: 1998.

1994–95 Teacher Follow-up Survey, Data File User's Manual Restricted-Use Codebook, NCES Working Paper 1999–14, by S.D. Whitener, K.J. Gruber, C.L. Rohr, and S.E. Fondelier. Washington, DC: 1999.

Data Quality and Comparability

Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey, NCES Working Paper 95–15, by J.E. Mullens. Washington, DC: 1995.

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, and M. Saba. Washington, DC: 1997.

Response Variance in the 1994–95 Teacher Follow-up Survey, NCES Working Paper 98–13, by J.M. Bushery, I.D. Schreiner, and A. Newman-Smith, Washington, DC: 1998.

The Results of the 1991–92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation, NCES Working Paper 95–10, by C.R. Jenkins and A. Wetzel. Washington, DC: 1995.

Chapter 6: National Education Longitudinal Study of 1988 (NELS:88)

1. OVERVIEW

The National Education Longitudinal Study of 1988 (NELS:88) is the third major secondary education longitudinal survey sponsored by NCES. The first two surveys—the National Longitudinal Study of the High School Class of 1972 (NLS-72) and the High School and Beyond (HS&B) Study—examined the educational, vocational, and personal development of young people, beginning in high school. (See chapters 7 and 8 for descriptions of these studies.) NELS:88 provides new data about critical transitions experienced by students from 8th grade through high school and into postsecondary education or the workforce. It expands the knowledge base of the two previous studies by surveying adolescents at an earlier age and following them into the 21st century.

The NELS:88 base year survey included a national probability sample of 1,052 public and private 8th-grade schools, with almost 25,000 participating students across the United States. Three follow-up surveys were conducted at 2-year intervals from 1990 to 1994. During 1994 (third follow up), most sample members were 2 years out of high school. A fourth follow up was conducted in 2000. In addition to surveying and testing students, NELS:88 gathered information from the parents of students, teachers, school administrators, and high school transcripts.

LONGITUDINAL SAMPLE SURVEY OF THE 8th-GRADE CLASS OF 1988; BASE-YEAR SURVEY AND FOUR FOLLOW UPS THROUGH 2000

NELS:88 collected data from:

- Students and dropouts
- School administrators
- Teachers
- Parents
- High school transcripts
- High school course offerings
- High School Effectiveness Study

Purpose

To (1) provide trend data about critical transitions experienced by young people as they leave elementary school and progress through high school into postsecondary institutions or the workforce, and (2) provide data for trend comparisons with results of the NLS-72 and HS&B studies.

Components

NELS:88 has collected survey data from students, dropouts, parents, teachers, and school administrators. Supplementary information has been gathered from high school transcripts and course-offering data provided by the schools, a Base Year Ineligible Study, and a High School Effectiveness Study. The various components are described below.

Base Year Survey. The base year survey was conducted during the spring school term in 1988, and included the following:

Student Questionnaire (8th-Grade Questionnaire). Students were asked to fill out a questionnaire that included items on their home background, language use, family, opinions about themselves, plans for the future, job and chores, school life, schoolwork, and activities. Students also completed a series of curriculum-based cognitive tests in four achievement areas—reading, mathematics, science, and social studies (history/government).

Parent Questionnaire. One parent of each student completed a questionnaire requesting information about both parents' background and socioeconomic characteristics, aspirations for their children, family willingness to commit resources to their children's education, the home educational support system, and other family characteristics relevant to achievement.

Teacher Questionnaire. A teacher questionnaire was administered to selected 8th-grade teachers responsible for instructing sampled students in two of the four test subjects—mathematics, science, English, and social studies. The questionnaire collected information in three areas: teachers' perceptions of the sampled students' classroom performances and personal characteristics; curriculum content of areas taught; and teachers' background and activities. Two teachers responded for each student.

School Administrator Questionnaire. Completed by an official in the participating school, this questionnaire collected information about school, student, and teacher characteristics; school policies and practices; the school's grading and testing structure; school programs and facilities; parent involvement in the school; and school climate.

First Follow-up Survey. The first follow-up survey was conducted in spring 1990. It collected information from students, teachers, and school administrators, but not parents. The student sample was freshened to be nationally representative of students enrolled in the 10th grade in spring 1990. In addition, three new components were initiated: the Dropout Questionnaire, the Base Year Ineligible (BYI) Study, and the High School Effectiveness Study (HSES).

Students were again requested to complete a questionnaire and take cognitive tests. The Student Questionnaire collected background information and asked students about such topics as their school and home environments, participation in classes and extracurricular activities, current jobs, goals and aspirations, and opinions about themselves. Dropouts were asked similar questions in a separate *Not Currently In School Questionnaire* (or *Dropout Questionnaire)*, which also requested specific information about reason(s) for leaving school and experiences in and out of school. Dropouts were also given cognitive tests.

School administrators provided information about their high schools in the School Administrator Questionnaire, and two teachers for each student completed the Teacher Questionnaire. There were different Teacher Questionnaires for English, mathematics, science, and history. The School Administrator and Teacher Questionnaires provided information about school administration, school programs and services, curriculum and instruction, and teachers' perceptions about their students' learning.

Second Follow-up Survey. The second follow-up survey, conducted in 1992, repeated all components of the first follow-up study and reinstated the Parent Questionnaire. The student sample was again freshened to be nationally representative of students enrolled in the 12th grade in spring 1992. A new Transcript Study provided archival data on the academic experience of high school students. Students in high schools designated in the first follow up for HSES were surveyed and tested again in both the main second follow-up survey and a separate HSES survey.

As in the previous waves, students were asked to complete a questionnaire and cognitive tests. The cognitive tests were designed to measure 12th-grade achievement and cognitive growth between 1988 and 1992 in mathematics, science, reading, and social studies (history/citizenship/geography). The questionnaire asked students about such topics as academic achievement; perceptions about their curricula and schools; family structures and environments; social relations; and aspirations, attitudes, and values relating to high school, occupations, and postsecondary education. The Student Questionnaire also contained an *Early Graduate Supplement*, which asked early graduates to document the reasons for and circumstances of their early graduation. Students who were first-time participants in NELS:88 completed a *New Student Supplement*, containing basic demographic items requested in the base year but not repeated in the second follow up. First follow-up dropouts were resurveyed and retested. School administrators completed the School Administrator Questionnaire, and one mathematics or science teacher for each student completed the Teacher Questionnaire.

Third Follow-up Survey. The third follow-up survey, conducted in 1994, contained only the Student Questionnaire, which collected information on issues of employment and postsecondary education. Specific content areas included academic achievement; perceptions and feelings about school and/or job; work experience and work-related training; application and enrollment in postsecondary education institutions; sexual behavior, marriage, and family; and values, leisure time activities, volunteer activities, and voting behavior.

Fourth Follow-up Survey. The fourth follow-up survey, conducted in 2000, contained only the Student Questionnaire, which collected information on issues of employment and postsecondary education. Specific content areas included academic achievement; perceptions and feelings about school and/or job; work experience and work-related training; application and enrollment in postsecondary education institutions; sexual behavior, marriage, and family; and values, leisure time activities, volunteer activities, and voting behavior.

Supplemental Studies. The following supplemental studies were conducted during the course of the NELS:88 project:

Base Year Ineligible (BYI) Study. The BYI Study was added to the first follow-up survey to ascertain the status of students who were excluded from the base year survey due to a language barrier or physical or mental disability that precluded them from completing a questionnaire and cognitive tests. Any students found to be eligible at this time were included in the follow-up surveys.

Followback Study of Excluded Students (FSES). This study—a part of the second follow-up survey—was a continuation of the first follow-up Base Year Ineligible Study.

Transcript Study. This study collected high school transcripts during the second follow-up survey. Complete transcript records were collected for (1) students attending sampled schools in spring 1992; (2) dropouts (including those in alternative programs) and early graduates; and (3) sample members who were ineligible for any wave of the survey due to mental or physical disability or language barriers.

High School Effectiveness Study (HSES). To facilitate longitudinal analysis at the school level, a School Effects Augmentation was implemented in the first follow-up survey to provide a valid probability sample of 10th-grade schools. From the pool of NELS:88 first follow-up schools, a probability subsample of 251 urban and suburban schools in the 30 largest Metropolitan Statistical Areas was selected for the HSES; 248 of these schools were final HSES participants in the first follow up. The NELS:88 national or "core" student sample in these schools was augmented to obtain a within-school representative student sample large enough to support school effects research (e.g., the effects of school policies and practices on students). These schools and students were followed up in 1992—when the majority of the students were in 12th grade—as part of both the main NELS:88 second follow-up survey and the HSES survey. The HSES also provided a convenient framework for a constructed response testing experiment in 1992. The test contained four questions that required students to derive answers from their own knowledge and experience (e.g., write an explanation, draw a diagram, solve a problem). Mathematics tests were assigned to half of the schools that were willing to commit the extra time required for such testing; the other half were assigned science tests. The second follow-up HSES was also enhanced by the collection of curriculum offerings in the Course Offerings Component. (See below.)

Course Offerings Component. This component was added to the second follow up to provide curriculum data that can serve as a baseline for studying student outcomes. Course offerings were collected from the HSES schools. (See above.) These data illuminate trends when contrasted to the transcript studies conducted as part of the 1982 HS&B and the 1987, 1990, 1994, and 1998 National Assessment of Educational Progress.

Periodicity

Biennial from 1988 to 1994. A fourth follow up was conducted in 2000. A Base Year Ineligible Study was conducted in 1990 as part of the first follow up; a continuation study, the Followback Study of Excluded Students, was conducted in 1992 as part of the second follow up. A High School Effectiveness Study was conducted in the first and second follow ups. A Transcript Study was implemented in the second follow up.

2. USES OF DATA

The NELS:88 project was designed to provide trend data about critical transitions experienced by students as they leave elementary school and progress through high school and into postsecondary education or the workforce. Its longitudinal design permits the examination of changes in young people's lives and the role of school in promoting growth and positive life outcomes. The project collects policy-relevant data about educational processes and outcomes, early and late predictors of dropping out, and school effects on students' access to programs and equal opportunity to learn. These data complement and strengthen state and local efforts by furnishing new information on how school policies, teacher practices, and family involvement affect student educational outcomes (e.g., academic achievement, persistence in school, and participation in postsecondary education).

NELS:88 data can be analyzed in three ways: cross-wave, cross-sectional, and cross-cohort (by comparing NELS:88 findings with those of the NLS-72 and HS&B studies). By following young adolescents at an earlier age (8th grade) and into the 21st century, NELS:88 expands the base of knowledge established in the NLS-72 and HS&B studies. NELS:88 first follow-up data provide a comparison point to high school sophomores 10 years earlier, as studied in HS&B. Second follow-up data allow trend comparisons of the high school class of 1992 with the 1972 and 1980 seniors studied in the NLS-72 and HS&B studies, respectively. The third follow up allows comparisons with NLS-72 and HS&B related to postsecondary outcomes. The three studies together provide measures of educational attainment in the United States and rich resources for studying the reasons for and consequences of academic success and failure.

More specifically, NELS:88 data can be used to investigate:

- *transitions from elementary to secondary school:* how students are assigned to curricular programs and courses; how such assignments affect their academic performance as well as future career and postsecondary education choices;
- *academic growth over time:* family, community, school, and classroom factors that promote growth; school classroom characteristics and practices that promote learning; effects of changing family composition on academic growth;
- *features of effective schools:* school attributes associated with student academic achievement; school effects analyses;
- *dropout process:* contextual factors associated with dropping out; movement in and out of school, including alternative high school programs;
- *role of the school in helping the disadvantaged:* school experiences of the disadvantaged; approaches that hold the greatest potential for helping them;
- *school experiences and academic performance of language minority students:* variation in achievement levels; bilingual education needs and experiences;
- *attracting students to mathematics and science:* math and science preparation received by students; student interest in these subjects; encouragement by teachers and school to study advanced mathematics and science; and
- *transitions from high school to college and postsecondary access/choice:* planning and application behaviors of the high school class of 1992; subsequent enrollment in postsecondary institutions.

3. KEY CONCEPTS

Some of the key terms related to NELS are defined below.

Cognitive Test Battery. Cognitive tests measuring student achievement in mathematics, reading, science, and social studies (history/citizenship/geography) were administered in the base year, first follow up, and second follow up. The contents was as follows: (1) reading (21 items, 21 minutes); (2) mathematics (40 items, 30 minutes); (3) sciences (25 items, 20 minutes); and (4) social studies (30 items, 14 minutes—the base year test included history and government items, the first and second follow-up tests included history, citizenship, and geography items).

Socioeconomic Status (SES). A composite variable constructed from five questions on the Parent Questionnaire: father's education level, mother's education level, father's occupation, mother's occupation, and family income. When all parent variables were missing, student data were used to compute socioeconomic status, substituting household items (e.g., dictionary, computer, more than 50 books, washing machine, calculator) for the family income variable. There are separate SES variables derived from parent data in the base year and the second follow up. The database also included variables for SES quartiles.

Dropout. Used both to describe an event (leaving school before graduating) and a status (an individual who was not in school and not a graduate at a defined point in time). The NELS:88 "cohort dropout rate" is based on a measurement of the enrollment status of 1988 8th graders 2 and 4 years later (in spring 1990 and spring 1992) and of 1990 sophomores 2 years later (in spring 1992). For a given point in time, a respondent is considered to be a dropout if he/she had not graduated from high school or attained an equivalency certificate and had not attended high school for 20 consecutive days (not counting excused absences). Transferring to another school is not regarded as a dropout event, nor is delayed graduation if a student was continuously enrolled but took an additional year to complete high school. A person who dropped out of school may have returned later and graduated. This person would be considered a "dropout" at the time he/she initially left school and a "stopout" at the time he/she returned to school.

4. SURVEY DESIGN

Target Population

Students enrolled in the 8th grade in "regular" public and private schools located in the 50 states and the District of Columbia during the spring 1988 school term. The sample was freshened in both the first and second follow ups to provide valid probability samples that would be nationally representative of 10th graders in spring 1990 and 12th graders in spring 1992. The NELS:88 project excludes the following types of schools: Bureau of Indian Affairs schools, special education schools for the handicapped, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel overseas. The following students are also excluded: mentally handicapped students and students not proficient in English, for whom the NELS:88 tests would be unsuitable; and students having physical or emotional problems that would make participation in the survey unwise or unduly difficult. However, a Base Year Ineligible Study (in the first follow up) and a Followback Study of Excluded Students (in the second follow up) sampled excluded students and added those no longer considered ineligible to the freshened sample of the first and second follow ups, respectively.

Sample Design

NELS:88 was designed to follow a nationally representative longitudinal component of students who were in the 8th grade in spring 1988. It also provides a nationally representative sample of schools offering 8th grade in 1988. In addition, by freshening the student sample in the first and second follow ups, NELS:88 provides nationally representative populations of 10th graders in 1990 and 12th graders in 1992. To meet the needs for cross-sectional, longitudinal, and cross-cohort analyses, NELS:88 involved complex research designs, including both longitudinal and cross-sectional sample designs.

Base Year Survey. In the base year, students were selected using a two-stage stratified probability design, with schools as the first-stage units and students within schools as the second-stage units. From a national frame of about 39,000 schools with 8th grades, a pool of 1,032 schools was selected through stratified sampling with probability of selection proportional to their estimated 8th-grade enrollment; private schools were oversampled to assure adequate representation. A pool of 1,032 replacement schools was selected by the same method to be used as substitutions for ineligible or refusal schools in the initial pool. A total of 1,057 schools cooperated in the base year; of these, 1,052 schools (815 public and 237 private) contributed usable student data. The sampling frame for NELS:88 was the school database compiled by Quality Education Data, Inc. of Denver, Colorado, supplemented by racial/ethnic data obtained from the U.S. Office for Civil Rights and school district personnel.

Student sampling produced a random selection of 26,435 8th graders in 1988; 24,599 participated in the base year survey. Hispanic and Asian/Pacific Islander students were oversampled. Within each school, approximately 26 students were randomly selected (typically, 24 regularly sampled students and 2 oversampled Hispanic or Asian/Pacific Islander students). In schools with fewer than 24 8th graders, all eligible students were selected. Potential sample members were considered ineligible and excluded from the survey if disabilities or language barriers were seen as obstacles to successful completion of the survey. The eligibility status of excluded members was reassessed in the first and second follow ups. (See below.)

First Follow-up Survey. There were three basic objectives for the first follow-up sample design. First, the sample was to include approximately 21,500 students who were in the 8th-grade sample in 1988 (including base year nonrespondents), distributed across 1,500 schools. Second, the sample was to constitute a valid probability sample of all students enrolled in the 10th grade in spring 1990. This entailed "freshening" the sample with students who were 10th graders in 1990 but who were not in the 8th grade in spring 1988 or who were out of the country at the time of base-year sampling. The freshening procedure added 1,229 10th graders; 1,043 of this new group were found to be eligible and were retained after final subsampling for the first follow-up survey. Third, the first follow up was to include a sample of students who had been deemed ineligible for base-year data collection due to physical, mental, or linguistic barriers to participation. The Base Year Ineligible Study reassessed the eligibility of these students so that those able to take part in the survey could be added to the first follow-up student sample. Demographic and school enrollment information was also collected for all students excluded in the base year, regardless of their eligibility status for the first follow up.

While schools covered in the NELS:88 base year survey were representative of the national population of schools offering the 8th grade, the schools in the first follow up were not representative of the national population of high schools offering the 10th grade. By 1990, the 1988 8th graders had dispersed to many high schools, which did not constitute a national probability sample of high schools. To compensate for this limitation, HSES was designed to sustain analyses of school effectiveness issues; HSES was conducted in conjunction with the first follow up. From the pool of participating first follow-up schools, a probability subsample of 251 urban and suburban schools in the 30 largest Metropolitan Statistical Areas was designated as HSES schools. The NELS:88 core student sample was augmented to obtain a within-school representative student sample large enough to support school effects research. The student sample was increased in HSES schools by an average of 15 students to obtain within-school student cluster sizes of approximately 30 students.

Second Follow-up Survey. The second follow-up sample included all students and dropouts selected in the first follow up. From within the schools attended by the sample members, 1,500 12th-grade schools were selected as sampled schools. Of these, the full complement of component activities occurred in 1,374 schools. For students attending schools other than those 1,374 schools, only the Student and Parent Questionnaires were administered. As in the first follow up, the student sample was augmented through freshening to provide a representative sample of students enrolled in the 12th grade in spring 1992. Freshening added into the sample 243 eligible 12th graders who were not in either the base year or first follow-up sampling frames. Schools and students designated for the HSES in the first follow up were followed up again—as part of both the NELS:88 second follow-up national survey and the HSES survey. The Followback Study of Excluded Students was a continuation of the first follow-up Base Year Ineligible Study. In addition, two new components—the Transcript Study and the Course Offerings Component—were added to the second follow up.

Third Follow-up Survey. The third follow-up student sample was created by dividing the second follow-up sample into 18 groups based on students' response history, dropout status, eligibility status, school sector type, race, test scores, socioeconomic status, and freshened status. Each sampling group was assigned an overall selection probability. Cases within a group were selected such that the overall group probability was met, but the probability of selection within the group was proportional to each sample member's second follow-up design weight. Assigning selection probabilities in this way reduced the variability of the third follow-up raw weights and consequently increased the efficiency of the resulting sample from 40.1 percent to 44.0 percent.

Fourth Follow-up Survey. The fourth follow-up student sample was the same as the third follow-up student sample.

Data Collection and Processing

NELS:88 compiled data from five primary sources: students, parents, school administrators, teachers, and high school administrative records (transcripts, course offerings, and course enrollments). Data collection efforts for the base year through third follow up extended from spring 1988 to summer 1994. Self-administered questionnaires, cognitive tests, and telephone or personal interviews were used to collect the data. The follow-up surveys involved extensive efforts to locate and collect data from sample members who were school dropouts, school transfers, or otherwise mobile individuals. Coding and editing conventions adhered as closely as possible to the procedures and standards previously established for the NLS-72 and HS&B. The National Opinion Research Center (NORC) at the University of Chicago was the prime contractor for the NELS:88 project from base year through the third follow up, but Research Triangle Institute conducted the fourth follow up.

Reference dates. In the base year survey, most questions referred to the student's experience up to the time of administration in spring 1988. In the follow ups, most questions referred to experiences that occurred between the previous survey and the current survey. For example, the second follow up largely covered the period between 1990 (when the first follow up was conducted) and 1992 (when the second follow up was conducted).

Data collection. Prior to each survey, it was necessary to secure a commitment to participate in the study from the administrator of each sampled school. For public schools, the process began by contacting the Council of Chief State School Officers and the officer in each state. Once approval was gained at the state level, contact was made with District Superintendents and then with school principals. For private schools, the National Catholic Educational Association and the National Association of Independent Schools were contacted for endorsement of the project, followed by contact of the school principals. The principal of each cooperating school designated a School Coordinator to serve as a liaison between NORC staff and selected respondents—students, parents, teachers, and the school administrator. The School Coordinator (most often a guidance counselor or senior teacher) handled all requests for data and materials, as well as all logistical arrangements for student-level data collection on the school premises. Coordinators were asked to identify students whose physical or learning disabilities or linguistic deficiencies would preclude participation in the survey and to classify all eligible students as Hispanic, Asian-Pacific Islander, or "other" race.

For the base year through second follow-up surveys, Student Questionnaires and test batteries were primarily administered in group sessions at the schools on a scheduled Survey Day. The sessions were monitored by NORC field staff, who also checked the questionnaires for missing data and attempted data retrieval while the students were in the classroom. Makeup sessions were scheduled for students who were unable to attend the first session. In the first and second follow ups, off-campus sessions were used for dropouts and for sample members who were not enrolled in a first follow-up school on Survey Day. The School Administrator, Teacher, and Parent Questionnaires were self-administered. NORC followed up by telephone with individuals who had not returned their questionnaires by mail within a reasonable amount of time.

The first follow-up data collection required intensive tracing efforts to locate base-year sample members who, by 1990, were no longer in their 8th-grade schools but had dispersed to many high schools. Also, in order to derive a more precise dropout rate for the 1988 8th-grade cohort, a second data collection was undertaken 1 year later, in spring 1991. At this time, an attempt was made to administer questionnaires—by telephone or in person—to sample members who had missed data collection at their school or who were no longer enrolled in school. The first follow up also included a Base Year Ineligible (BYI) Study, which surveyed a sample of students considered ineligible in the base year due to linguistic, mental, or physical deficiencies. The BYI Study sought to determine if eligibility status had changed for the excluded students so that newly eligible students could be added to the longitudinal sample. If an excluded student was now eligible, an abbreviated Student Questionnaire or a Dropout Questionnaire was administered, as appropriate. For those students who were still ineligible, their school enrollment status was ascertained and basic information about their sociodemographic characteristics was recorded.

Tracing efforts continued in the second and third follow ups. In the second follow up (conducted in 1992), previously excluded students were surveyed through the Followback Study of Excluded Students. The second follow up also collected transcripts, course offerings, and course enrollments from the high schools; reminder postcards were sent to principals who did not respond within a reasonable period. Data collection for HSES was conducted concurrently with the collection for the second follow up. Because of the overlap in school and student samples, survey instruments and procedures for HSES were almost identical to those used in the main NELS:88 survey.

By 1994, when the third follow up was conducted, most sample members had graduated from high school and it was no longer feasible to use group sessions to administer Student Questionnaires. Instead, the dominant form of data collection was one-on-one administration through computer-assisted telephone interviewing (CATI). In-person interviews were used for sample members who required intensive in-person locating or refusal conversion. Only the Student Questionnaire was administered in the third follow up.

By 2000, when the fourth follow up was conducted, most sample members who attended college and technical schools had completed their postsecondary education. The survey was conducted primarily by computer-assisted telephone interviewing.

Processing. Data processing activities were quite similar for the base year survey and the first and second follow ups. An initial check of student documents for missing data was performed on-site by NORC staff so that data could be retrieved from the students before they left the classroom. Special attention was paid to a list of "critical items." Once the questionnaires and tests were received at NORC, they were again reviewed for completeness, and a final disposition code was assigned to the case indicating which documents had been completed by the sample member. Postsecondary institutions reported by the student were coded using the standard Integrated Postsecondary Education Data System (IPEDS) codes. Data entry for both Student Questionnaires and cognitive tests was performed through optical scanning. New Student Supplements and Dropout Questionnaires were converted to machine-readable form using key-to-disk methods. All cognitive tests were photographed onto microfilm for archival storage.

In the third follow up, a CATI system captured the data at the time of the interview. The system evaluated the responses to completed questions and used the results to route the interviewer to the next appropriate question.

The CATI program also applied the customary edits, described below under "Editing." At the conclusion of an interview, the completed case was deposited in the database ready for analysis. There was minimal post-data entry cleaning because the interviewing module itself conducted the majority of necessary edit checking and conversion functions.

Verbatim responses were collected in the third follow up for a number of items, including occupation and major field of study. When respondents indicated their occupation, the CATI interviewers recorded the verbatim response. The system checked the response using a keyword search to match it to a subset of standard industry and occupation codes, and then presented the interviewer with a set of choices based on the keyword matches. The interviewer chose the option which most closely matched the information provided by the respondent, probing for additional information when necessary. Quality control was ensured by a reading and recoding, if necessary, of the verbatim responses by professional readers.

Editing. In the base year through second follow-up surveys, detection of out-of-range codes was completed during scanning or data entry for all closed-ended questions. Machine editing was used to: (1) resolve inconsistencies between filter and dependent questions; (2) supply appropriate missing data codes for questions left blank (e.g., legitimate skip, refusal); (3) detect illegal codes and convert them to missing data codes; and (4) investigate inconsistencies or contradictions. Frequencies and crosstabulations for each variable were inspected before and after these steps to verify the accuracy and appropriateness of the machine editing. Items with unusually high nonresponse or multiple responses were further checked by verifying the responses on the questionnaire. A final editing step involved recoding Student Questionnaire responses for some items to the codes for the same items in earlier NELS:88 waves or in HS&B. Once this was done, codes that differed on the Dropout Questionnaire were recoded to coincide with the codes used for Student Questionnaire responses.

In the third follow up, machine editing was replaced by the interactive edit capabilities of the CATI system, which tested responses for valid ranges, data field size, data type (numeric or text), and consistency with other answers or data from previous rounds. If the system detected an inconsistency because of an interviewer's incorrect entry, or if the respondent simply realized that he or she made a reporting error earlier in the interview, the interviewer could go back and change the earlier response. As the new response was entered, all of the edit checks performed at the first response were again performed. The system then worked its way forward through the questionnaire using the new value in all skip instructions, consistency checks, and the like until it reached the first unanswered question, and control was then returned to the interviewer. When problems were encountered, the system could suggest prompts for the interviewer to use in eliciting a better or more complete answer.

Estimation Methods

Sample weighting is required that NELS:88 data are representative. Imputation for missing nonresponses, however, has not yet been systematically provided for data analysis.

Weighting. Weighting is used in NELS:88 data analysis to accomplish a number of objectives, including: (1) to expand counts from sample data to full population levels; (2) to adjust for differential selection probabilities (e.g., the oversampling of Asian and Hispanic students); (3) to adjust for differential response rates; and (4) to improve representativeness by using auxiliary information. Multiple "final" (or nonresponse-adjusted) weights have been provided for analyzing the different populations that NELS:88 data represent (i.e., base year schools; 8th graders in 1988 and 2, 4, and 6 years later; 1990 sophomores; 1992 seniors). Weights should be used together with the appropriate flag in order to analyze the sample for a particular targeted population.

Weights have not been constructed for all possible analytic purposes. In cases where no specific weight is available, existing weights may provide reasonable approximations. For instance, base year parent and cognitive test completion rates were so high relative to student questionnaire completion that the student weight can be used for them with minimal bias.

NELS:88 weights were calculated in two steps: (1) unadjusted weights were calculated as the inverse of the probabilities of selection, taking into account all stages of the sample selection process; and (2) these initial weights were adjusted to compensate for nonresponse, typically carried out separately within multiple weighting cells. For detailed discussions of the calculation of weights for each wave, users are referred to the methodology reports for the individual surveys.

Scaling (item response theory). Item response theory (IRT) was used to calibrate item parameters for all cognitive test items administered to students in NELS:88

assessments. The tests conducted in each NELS:88 survey generated achievement measures in standardized scores, and grade 12 mathematics scores equivalent to those in the National Assessment of Educational Progress (NAEP) surveys, among others. For detail about IRT-based cognitive test design, see chapter 20.

Imputation. NELS:88 surveys have not involved large-scale imputation of missing data. Only a few variables have been imputed: student's sex, race/ethnicity, and school enrollment status. For example, when sex was missing in the data file, the information was looked for on earlier school rosters. If it was still unavailable after this review, sex was assumed from the sample member's name (if unambiguous). As a final resort, sex was randomly assigned.

5. DATA QUALITY AND COMPARABILITY

A number of studies have been conducted to address data quality issues relating to the NELS:88 project. During the course of data collection and processing, systematic efforts were made to monitor, assess, and maximize data quality. Subsequent studies were conducted to evaluate the data quality in comparison with earlier longitudinal surveys.

Sampling Error

Because the NELS:88 sample design involved stratification, disproportionate sampling of certain strata, and clustered (i.e., multistage) probability sampling, the calculation of exact standard errors (an indication of sampling error) for survey estimates can be difficult and expensive. NORC used the Taylor Series procedure to calculate the standard errors for NELS:88 estimates.

Standard errors and design effects for about 30 key variables in each NELS:88 wave from the base year through the second follow up were calculated using SUDAAN software. These can be used to approximate the standard errors if users do not have access to specialized software.

Design effects. A comparative study of design effects across NELS:88 waves and between NELS:88 and HS&B was done. When comparing NELS:88 base year student questionnaire data to the results from HS&B—the 30 variables from the NELS:88 student questionnaire were selected to overlap as much as possible with those variables examined in HS&B—the design effects indicate that the NELS:88 sample was slightly more efficient than HS&B. The smaller design effects in the NELS:88 base year may reflect its smaller cluster size (24 students plus, on average, two oversampled Hispanics and Asian from each NELS:88 school versus the 36 sophomore and 36 senior selections from each HS&B school). The mean design effect for base year students is 2.54.

In the comparative study of design effects across NELS:88 waves, the design effects in the first follow up were somewhat higher than those of the base year, a result of the subsampling procedures used for the first follow up. The mean design effect for 1st follow up students and dropouts is 3.80. The conditional design effects in the 2nd follow up are lower than those in the 1st follow up, but higher than those in the base year. The conditional mean design effect for 2nd follow up students and dropouts is 3.71. (See *NELS:88 Base Year Through Second Follow-up Final Methodology Report*, NCES Working Paper 98–06.)

Nonsampling Error

Coverage error. Exclusion and undercoverage of certain groups of schools and students in NELS:88 generated coverage error. In the base year survey, for example, students who had linguistic, mental, or physical obstacles were excluded from the study. Consequently, the national populations for such student groups were not fully covered by the sample.

To correct this coverage bias, a Base Year Ineligible (BYI) Study collected eligibility information for 93.9 percent of the sample members excluded in the base year survey. For those who were reclassified as eligible in the BYI Study, Student or Dropout Questionnaires were administered in person or over the telephone during the first follow up. Cognitive tests were also administered to a small percentage of these students. For students who remained ineligible, school enrollment status and other key characteristics were obtained. The BYI Study permitted an evaluation of coverage bias in NELS:88 and a means of reducing undercoverage by identifying newly eligible students who could then be added into the sample to ensure cross-sectional representativeness. This effort also provided a basis for making corrected dropout estimates, taking into account both 1988-eligible and 1988-ineligible 8th graders 2 years later. For details on the BYI Study, see *Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years* (NCES 96-723).

Nonresponse error. Both unit nonresponse (nonparticipation in the survey by a sample member) and item nonresponse (missing value for a given questionnaire/test item) have been evaluated in NELS:88 data.

Unit nonresponse. In the NELS:88 base year survey the initial school response rate was 69 percent. This low rate prompted a follow-up survey to collect basic characteristics from a sample of the nonparticipating schools. These data were then compared to the same characteristics among the participating schools to assess the possible impact of response bias on the survey estimates. The school-level nonresponse bias was found to be small to the extent that schools could be characterized by size, control, organizational structure, student composition, and other factors. Bias at the school level was not assessed for the follow-up surveys because (1) sampling for the first and second follow ups was student-driven (i.e., the schools were identified by following student sample members) and the third follow up did not involve schools; and (2) school cooperation rates were very high (up to 99 percent). Even if a school refused to cooperate, individual students were pursued outside of school (although school context data were not collected). The student response rates are shown in the table below.

Student-level nonresponse analysis was conducted with a focus on *panel nonresponse* since a priority of the NELS:88 project is to provide a basis for longitudinal analysis. Nonresponse was examined for the 8th-grade and 10th-grade cohorts. Any member of the 8th-grade cohort who did not complete a survey in three rounds (base year, first follow up, and second follow up) and any member in the 10th-grade cohort who did not complete a survey in the second and third rounds (first and second follow ups) was considered a panel nonrespondent for that cohort. Panel nonresponse to cognitive tests in the two cohorts were defined the same way. The nonresponse rate was defined as the proportion of the selected students (excluding deceased students) who were nonrespondents in any round in which data were collected.

Nonresponse rates for both cohorts were calculated by school- and student-level variables that were assumed to be stable across survey waves (e.g., sex and race). These variables allowed comparison between participants and nonparticipants even though the data for the latter were missing in some rounds. Estimates were made with both weighted and unweighted data. The weight used was the second follow-up raw panel weight (not available in the public release data set). About 18 percent of the 8th-grade cohort and 10 percent of the 10th-grade cohort were survey nonrespondents at one or more points in time. Approximately 43 percent of the 8th-grade cohort and 35 percent of the 10th-grade cohort did not complete one or more cognitive tests in their rounds of testing.

Nonresponse bias was calculated as the difference in the estimates between the respondent and all selected students. On the whole, the analysis revealed only small discrepancies between the two cohorts. Bias estimates were higher, however, for the 8th-grade cohort than for the 10th-grade cohort because of the 8th-grade cohort's more stringent definition of participation. The discrepancies between cognitive test completers and noncompleters were larger than between survey participants and nonparticipants; this pattern held for both

Table 4. Unit level and overall level weighted response rates for selected NELS:88 student populations

Population	Unit level weighted response rate				
	Base year 1st level	Base year 2nd level	1st follow up	2nd follow up	3rd follow up
Interviewed students	*63.7	93.4	91.1	91.0	90.9
Tested students	*63.7	90.2	94.1	76.6	†
Dropouts	*63.7	†	91.0	88.0	†
Tested dropouts	*63.7	†	48.6	41.7	†
	Overall level weighted response rate				
	Base year 1st level	Base year 2nd level	1st follow up	2nd follow up	3rd follow up
Interviewed students	*63.7	59.4	58.0	58.0	57.9
Tested students	*63.7	57.5	59.9	37.4	†
Dropouts	*63.7	†	58.0	56.1	†
Tested dropouts	*63.7	†	31.0	26.6	†

*Unweighted response rate
†Not applicable
SOURCE: Seastrom, Salvucci, Walter, and Shelton (forthcoming), *A Review of the Use of Response Rates at NCES.*

cohorts. In brief, the magnitude of the bias was generally small—few percentage estimates were off by as much as 2 percent in the 8th-grade cohort and 1 percent in the 10th-grade cohort. Such bias reflects the raw weight. The nonresponse-adjusted weight should correct for differences by race and sex to produce correct population estimates for each subgroup.

Further analysis was done using several other student and school variables. The results showed rather similar patterns of bias. When compared with estimates from HS&B, the student nonresponse bias estimates in NELS:88 were consistently lower. However, the two studies seem to share certain common patterns of nonresponse. For example, both studies generated comparatively higher nonresponse rates among students enrolled in schools in the West, Black students, students in vocational or technical programs, students in the lowest test quartile, and dropouts.

Item nonresponse. Item nonresponse was examined in base year though second follow-up data obtained from surveys of students, parents, and teachers. Differences emerged among student subgroups in the level of nonresponse to a wide range of items—from language background, family composition, and parents' education to perception of school safety. Nonresponse was often two to five times as great for one subgroup as for the other subgroups. High item nonresponse rates were associated with such attributes as not living with parents, having low socioeconomic status, being male, having poor reading skills, and being enrolled in a public school. Compared with parent nonresponse to items about college choice and occupational expectations, student nonresponse rates were generally lower. For items about student's language proficiency, classroom practices, and student's high school track, students had consistently lower nonresponse rates than observed among their teachers. See *NELS:88 Survey Item Evaluation Report* (NCES 97-052) for further detail.

Measurement error. NCES has conducted studies to evaluate measurement error in (1) student questionnaire data compared to parent and teacher data, and (2) student cognitive test data.

Parent-student convergence and teacher-student convergence. A study of measurement error in data from the base year through second follow-up surveys focused on the convergence of responses by students and parents and by students and teachers. (See *NELS:88 Survey Item Evaluation Report*, NCES 97-052.) Response convergence (or discrepancy) across respondent groups can be interpreted as an indication of measurement reliability, validity, and communality, although data are often not sufficient to determine which response is more accurate.

The student and parent components of this study covered such variables as sibling size, student's work experience, language background, parents' education, parent-student discussion of issues, perceptions about school, and college and occupation expectations. Parent-student convergence varied from very high to very low, depending on the item. For example, convergence was high for the number of siblings, regardless of student-level characteristics such as socioeconomic status, sex, reading scores, public versus private school enrollment, and whether or not living with parents. In contrast, parent-student convergence was low for items related to the student's work experience; there was also more variation across student subgroups for these items. In general, convergence tended to be high for objective items, for items worded similarly, and for nonsensitive items.

Teacher-student convergence was examined through variables about student's English proficiency, classroom practices, and student's high school track. Again, convergence was found to vary considerably across data items and student subgroups. Convergence was high for student's native language but low for student's English proficiency. Across student subgroups, there was a greater range in the correlations for English proficiency than for native language. Teachers and students differed quite dramatically on items about classroom practices.

Cognitive test data. In-depth studies of measurement error issues related to cognitive tests administered in the base year through second follow-up surveys are also available. (See *Psychometric Report for the NELS:88 Base Year Test Battery*, NCES 91-468, and *Psychometric Report for the NELS:88 Base Year Through Second Follow-up*, NCES 95-382.)

The first study addressed issues related to test speediness (the limited testing time in relation to the outcome), reliability, item statistics, performance by racial/ethnic and gender groups, and Item Response Theory (IRT) parameters for the battery. The results indicate that the test battery either met or exceeded all of its psychometric objectives. Specifically, the following findings were reported: (1) while the allotted testing time was only 1½ hours, quite acceptable reliabilities were obtained for the tests on reading comprehension, mathematics, history/citizenship/geography, and, to a somewhat lesser extent, science; (2) the internal consistency reliabilities were sufficiently high to justify the use of IRT scoring, and thus provide the framework for constructing 10th- and

12th-grade forms that would be adaptive to the ability levels of the students; (3) there was no consistent evidence of differential item functioning (item bias) for gender or racial/ethnic groups; (4) factor analysis results supported the discriminant validity of the four tested content areas; convergent validity was also indicated by salient loadings of testlets composed of "marker items" on their hypothesized factors; and (5) in addition to providing the usual normative scores in all four tested areas, behaviorally anchored proficiency scores were provided in both the reading and math areas.

The second study focused on issues relating to the measurement of gain scores. Special procedures were designed into the test battery design and administration to minimize the floor and ceiling effects that typically distort gain scores. The battery used a two-stage multilevel procedure that attempted to tailor the difficulty of the test items to the performance level of a particular student. Thus, students who performed very well on their 8th-grade mathematics test received a relatively more difficult form in 10th grade than students who had not performed well on their 8th-grade test. There were three forms of varying difficulty in mathematics and two in reading in both grades 10 and 12. Since 10th and 12th graders were taking forms that were more appropriate for their level of ability/achievement, measurement accuracy was enhanced and floor and ceiling effects could be minimized. The remaining two content areas—science and history/citizenship/geography—were only designed to be grade-level adaptive (i.e., a different form for each grade but not multiple forms varying in difficulty within grade).

To maximize the gain from using an adaptive procedure, special vertical scaling procedures were used that allow for Bayesian priors on subpopulations for both item parameters and scale scores. In comparing more traditional non-Bayesian approaches to scaling longitudinal measures with the Bayesian approach, it was found that the multilevel approach did increase the accuracy of the measurement. Further, when used in combination with the Bayesian item parameter estimation, the multilevel approach reduced floor and ceiling effects when compared to the more traditional item response theory approaches.

Data Comparability

NELS:88 is designed to facilitate both longitudinal and trend analyses. Longitudinal analysis calls for data compatibility across survey waves whereas trend analysis requires data compatibility with other longitudinal surveys. Data compatibility issues may relate to survey instruments, sample design, and data collection methods.

Comparability within NELS:88 across survey waves. A large number of variables are common across survey waves. (See *NELS:88 Second Follow-up Student Component Data File User's Manual* for a listing of common Student Questionnaire variables in the base year, first follow up, and second follow up.) However, compatibility of NELS:88 data across waves can still be an issue because of subtle differences in question wording, sample differences (e.g., with or without dropouts and freshening students, sample attrition, nonresponse) and data collection methods (e.g., on-campus group session, off-campus individual survey, telephone interview).

One NCES study compared 112 pairs of variables repeated from the base year to the first and second follow-up surveys. (See *NELS:88 Survey Item Evaluation Report*, NCES 97-052.) These variables cover student family, attitudes, education plans, and perceptions about schools. The results suggest that the interpretations of NELS:88 items depend on the age level at which they were administered. Data convergence tended to be higher for pairs of first and second follow-up measures than for pairs of base year and second follow-up measures. Some measures were more stable than others. Students responded nearly identically to the base year and second follow-up questions about whether English was their native language. Their responses across survey waves were also fairly stable as to whether their curriculum was intended to prepare them for college, whether they planned to go to college, and their religiosity. It should be noted that cross-wave discrepancies may reflect a change in actual student behavior rather than a change in response for a status quo situation.

Comparability within NELS:88 across respondent groups. While different questionnaires were used to collect data from different respondent groups (students, parents, teachers, school administrators), there are overlapping items among these instruments. One study examined the extent to which the identical or similar items in different questionnaires generated compatible information. It found considerable discrepancies between students and parents, and even greater discrepancies between students and teachers, in their responses to selected groups of overlapping variables. (See earlier section on "Measurement error.")

Comparability with NLS-72 and HS&B. NELS:88 surveys contain many items that were also covered in

NLS-72 and HS&B—a feature that enables trend analyses of various designs. (See *NELS:88 Second Follow-up Student Component Data File User's Manual* for a crosswalk of common variables and a discussion of trend analyses.) To examine data compatibility across the three studies, one should consider their sample designs and data contents, including questionnaires, cognitive tests, and transcript records.

Sample designs for the three studies are similar. In each base year, students were selected through a two-stage stratified probability sample, with schools as the first-stage units and students within schools as the second-stage units. In NLS-72, all baseline sample members were spring term 1972 high school seniors. In HS&B, all members of the student sample were spring term 1980 sophomores or seniors. Because NELS:88 base year sample members were 8th graders in 1988, its follow ups encompass students (both in the modal grade progression sequence and out of sequence) and dropouts. Sample freshening was used in NELS:88 to provide cross-sectional nationally representative samples. Despite similarities, however, the sample designs of the three studies differ in three major ways: (1) the NELS:88 first and second follow ups had relatively variable, small, and unrepresentative within-school student samples, compared to the relatively uniform, large, and representative within-school student samples in the NLS-72 and HS&B studies; (2) unlike the two earlier projects, NELS:88 did not provide a nationally representative school sample in its follow ups; and (3) there were differences in school and subgroup sampling and oversampling strategies in the three studies. These sample differences imply differences in respondent populations covered by the three studies.

Questionnaire overlap is apparent among the three studies but, nevertheless, requires caution when making trend comparisons. Some items were repeated in identical form across the studies; others appear to be essentially similar but have small differences in wording or response categories.

Item response theory (IRT) was used in the three studies to put math, vocabulary, and reading test scores on the same scale for 1972, 1980, and 1982 seniors. Additionally, there were common items in the HS&B and NELS:88 math tests that provide a basis for equating 1980–1990 and 1982–1992 math results. In general, however, the tests in the three studies differed in many ways. *Although group differences by standard deviation units may profitably be examined, caution should be exercised in drawing time-lag comparisons for cognitive test data.*

Transcript studies in NELS:88, HS&B, and the National Assessment of Educational Progress (NAEP) were designed to support cross-cohort comparisons. The NAEP and NELS:88 studies, however, provide summary data in Carnegie units, whereas the HS&B provides course totals. Note too that course offerings were only collected for schools that were part of the High School Effectiveness Study in the NELS:88 second follow up whereas course offerings were collected for *all* schools in HS&B. (See chapter 8.)

Other factors should be considered in assessing data compatibility. Differences in mode and time of survey administration across the cohorts may affect compatibility. NELS:88 seniors were generally surveyed earlier in the school year than were NLS-72 seniors. NLS-72 survey forms were administered by school personnel while HS&B and NELS:88 survey forms were administered primarily by contractor staff. There were also differences in questionnaire formats; the later tests had improved mapping and different answer sheets.

6. CONTACT INFORMATION

For content information on the NELS:88 project, contact:

Jeffrey Owings
Phone: (202) 502–7423
E-mail: jeffrey.owings@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

National Education Longitudinal Study (NELS:88/94) Methodology Report, NCES 96–174, by C. Haggerty, B. Dugoni, L. Reed, A. Cederlund, and J. Taylor. Washington, DC: 1996.

NELS:88 Base Year Through Second Follow-Up: Final Methodology Report, NCES Working Paper 98-06, by S.J. Ingels, L.A. Scott, J.R. Taylor, J. Owings, and P. Quinn. Washington, DC: 1998.

NELS:88 Second Follow-Up: Dropout Component Data File User's Manual, NCES 93–375, by S.J. Ingels, K.L. Dowd, J.L Stipe, J.D. Baldridge, V.H. Bartot, and M.R. Frankel. Washington, DC: 1995.

Procedures Guide for Transcript Studies, NCES Working Paper 1999–05, by M.N. Alt and D. Bradby. Washington, DC: 1999.

A Profile of Parents of Eighth Graders, NCES 90–488, by L. Horn and J. West. Washington, DC: 1992.

Uses of Data

A Guide to Using NELS:88 Data, by J. Owings, M. McMillen, S. Ahmed, J. West, P. Quinn, E. Hausken, R. Lee, S. Ingels, L. Scott, D. Rock, and J. Pollack. Washington, DC: 1994. (Prepared for the 1994 AERA Annual Meeting in New Orleans, LA)

National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data, NCES Working Paper 95–06, by S. Ingels and J. Taylor. Washington, DC: 1995.

National Education Longitudinal Study of 1988: Conducting Trend Analyses: HS&B and NELS:88 Dropouts, NCES Working Paper 95–07, by S. Ingels and K. Dowd. Washington, DC: 1995.

National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors, NCES Working Paper 95–05, by S. Ingels and J. Baldridge. Washington, DC: 1995.

Survey Design

NELS:88 Base Year Sample Design Report, NCES 90–463, by B.D. Spencer, M.R. Frankel, S.J. Ingels, K.A. Rasinski, and R.E. Tourangeau. Washington, DC: 1990.

NELS:88 Second Follow-up Research and Development Working Papers, NCES 94–251, by S. Ingels. Washington, DC: 1995.

1998 Revision of the Secondary School Taxonomy, NCES Working Paper 1999–06, by D. Bradby and G. Hoachlander. Washington, DC: 1999.

Psychometric Report for the NELS:88 Base Year Test Battery, NCES 91–468, by D.A. Rock, J.M. Pollack, J. Owings, and A. Hafner. Washington, DC: 1991.

Sample Exclusion in NELS:88: Characteristics of Base Year Ineligible Students; Changes in Eligibility Status After Four Years, NCES 96–723, by S.J. Ingels. Washington, DC: 1996.

Data Quality and Comparability

Constructed Response Tests in the NELS:88 High School Effectiveness Study, NCES Working Paper No. 97–804, by J.M. Pollock, and D.A. Rock. Washington, DC: 1997.

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, and M. Saba. Washington, DC: 1997.

NELS:88 Survey Item Evaluation Report, NCES 97–052, by D.H. McLaughlin, J. Cohen, and R. Lee. Washington, DC: 1997.

Psychometric Report for the NELS:88 Base Year Through Second Follow-up, NCES 95–382, by D.A. Rock, J.M. Pollack, and P. Quinn. Washington, DC: 1995.

Quality of Responses of Eighth-Grade Students in NELS:88, NCES 90–487, by P. Kaufman and K.A. Rasinski. Washington, DC: 1991.

A Review of the Use of Response Rates at NCES, forthcoming, by M. Seastrom, S. Salvucci, E. Walter, and K. Shelton.

Strengths and Limitations for Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets, NCES Working Paper 2000–03, by P. Broene and K. Rust. Washington, DC: 2000.

Chapter 7: National Longitudinal Study of the High School Class of 1972 (NLS-72)

1. OVERVIEW

LONGITUDINAL SAMPLE SURVEY OF THE HIGH SCHOOL SENIOR CLASS OF 1972. BASE-YEAR SURVEY AND FIVE FOLLOW UPS, ENDING IN 1986

NLS-72 collected data from:
- Students
- School administrators
- School counselors
- Postsecondary transcripts

In response to the need for policy-relevant, time-series data on nationally representative samples of elementary and secondary students, NCES instituted the National Longitudinal Studies Program, a continuing long-term project. The general aim of this program is to study the educational, vocational, and personal development of students at various grade levels, and the personal, familial, social, institutional, and cultural factors that may affect that development. The National Longitudinal Study of the High School Class of 1972 (NLS-72) was the first in the series. The first three studies—NLS-72, the High School and Beyond Study (see chapter 8), and the National Education Longitudinal Study of 1988 (see chapter 6)—cover the educational experience of youth from the 1970s into the 1990s.

NLS-72 collected comprehensive base-year data from a nationally representative sample of high school seniors in spring 1972, prior to high school graduation. Additional information about students and schools was obtained from school administrators and counselors. Over the course of the project—extending from the base-year survey in 1972 to the fifth follow-up survey in 1986—data were collected on nearly 23,000 students. A number of supplemental data collection efforts were also undertaken, including a Postsecondary Education Transcript Study (PETS) in 1984–85, and a Teaching Supplement in 1986.

Purpose

To provide information on the transitions of young adults from high school through postsecondary education and into the workplace.

Components

NLS-72 collected data from students (seniors in 1972), school administrators, and school counselors. Data were primarily collected in a base-year and five follow-up surveys. The project also included periodic supplements completed by 1972 seniors and a collection of postsecondary transcripts from colleges and universities attended by the students.

Base-Year Survey. The base-year survey was conducted in spring 1972 and comprised the following:

Student Questionnaire. Students reported information about their personal and family background (age, sex, race, physical handicap, socioeconomic status of family and community); education and work experiences (school characteristics and performance, work status, performance and satisfaction); future plans (work, education, and/or military); and aspirations, attitudes, and opinions. Students also completed a *Test Battery*—six timed aptitude tests which measured verbal and nonverbal abilities. These tests covered vocabulary, picture number (two parts), reading, letter groups, mathematics, and mosaic comparisons (three parts).

Student Record Information Form (SRIF). School administrators completed this form for each student sample member. The SRIF collected data on each student's high school curriculum, credit hours in major courses, grade point average, and (if applicable) his or her position in ability groupings, remedial-instruction record, involvement in certain federally supported programs, and scores on standardized tests.

School Questionnaire. School administrators provided data on program and student enrollment information, such as grades covered, enrollment by grades, curricula offered, attendance records, racial/ethnic composition of school, dropout rates by sex, number of handicapped and disadvantaged students, and percentage of recent graduates in college.

Counselor Questionnaire. One or two counselors in each school provided data on their sex, race, and age; college courses in counseling and practice background; total years of counseling and years at present school; prior counseling experience with racial/ethnic minority groups; sources of support for postsecondary education recommended to/used by students; job placement methods used; number of students assigned for counseling and number counseled per week; time spent in counseling per week; time spent with students about various problems, choices, and guidance; and time spent in various other activities (e.g., conferences with parents and teachers).

Follow-up Surveys. In 1973, 1974, 1976, 1979, and 1986, NCES conducted follow-up surveys of students in the 1972 base-year sample and of students in an augmented sample selected for the first follow up. These surveys collected information from the 1972 seniors on marital status; children; community characteristics; education, military service, and/or work plans; educational attainment (schools attended, grades received, credits earned, financial assistance); work history; attitudes and opinions relating to self-esteem, goals, job satisfaction, and satisfaction with school experiences; and participation in community affairs or political activities. School Questionnaires and retrospective high school data were collected during the first follow up for sample schools and students who had not participated in the base-year survey.

Concurrently with the second follow up, an *Activity State Questionnaire* was administered to sample members who had not provided this information in the base-year or first follow-up surveys. Data were collected on pursuits in which the sample member was active in October of 1972 and 1973, including education, work, military service, being a housewife, and other activities. Background information about the sample member's high school program and about parents' education and occupation was also requested.

During the fourth follow-up survey, a subsample of sample members was retested on a subset of the base-year Test Battery. In addition, a *Supplemental Questionnaire* was administered to respondents who had not reported certain information in previous surveys. The information asked for retrospectively covered the sample member's school and employment status in October 1972 to 1976 and his/her license or diploma status as of October 1976. The questionnaires were tailored to the sample member's pattern of missing responses and consisted of two to four of the possible sections.

The fifth follow-up survey offered the opportunity to gather information on experiences and attitudes of a sample for whom an extensive history already existed. It differed from the previous follow ups in that it was only sent to a subsample of the original respondents and targeted certain subgroups in the population. About 10 pages of new questions on marital history, divorce, child support, and economic relationships in families were included. The fifth follow up also included a sequence of questions aimed at understanding the kinds of individuals who apply for and enroll in graduate management programs, as well as several questions about attitudes toward the teaching profession.

A *Teaching Supplement* was administered concurrently with the fifth follow up. A separate questionnaire was sent to fifth follow-up respondents who indicated on the main survey form that they had teaching experiences or had been trained for teaching. The instrument focused on the qualifications, experiences, and attitudes of current and former elementary and secondary school teachers, and on the qualifications of persons who had completed a degree in education or who had received certification but had not actually taught. Items included reasons for entering the teaching career, degrees and certification, actual teaching experience, allocation of time while working, pay scale, satisfaction with teaching, characteristics of the school in which the respondent taught, and professional activities. Former teachers were asked about their

BEST COPY AVAILABLE

reasons for leaving the teaching profession and the career (if any) they pursued afterward. Current teachers were asked about their future career plans, including how long they expected to remain in teaching. The supplement included six critical items: type of certification, certification subject(s), first year of teaching, beginning salary of the district where the respondent was currently teaching, years of experience, and the grade level taught.

Postsecondary Education Transcript Study (PETS). To provide data on coursework and credits for analysis of occupational and career outcomes, NCES requested official transcripts from all academic and vocational schools attended by the 1972 seniors since leaving high school. This study, conducted during 1984–85, collected transcripts from all postsecondary institutions reported by sample members in the first through fourth follow-up surveys. Information from transcripts include terms of attendance, fields of study, specific courses taken, and grades and credits earned. As the study covered a 12-year period, dates of attendance and term dates were recorded from each transcript received, allowing analysis over the whole period or any defined part.

Periodicity

The base-year survey was conducted in the spring of 1972, with five follow ups in 1973, 1974, 1976, 1979, and 1986. Supplemental data collections were administered during all but the third follow up. Postsecondary transcripts were collected in 1984–85.

2. USES OF DATA

NLS-72 is the oldest of the longitudinal studies sponsored by NCES. It is probably the richest archive ever assembled on a single generation of Americans. Young people's success in making the transition from high school or college to the workforce varies enormously for reasons only partially understood. NLS-72 data can provide information about quality, equity, and diversity of educational opportunity and the effect of those factors on cognitive growth, individual development, and educational outcomes. It can also provide information about changes in educational and career outcomes and other transitions over time.

The Teaching Supplement data can be used to investigate policy issues related to teacher quality and retention. These data can be linked to data from prior waves of the Student Questionnaire for analysis of antecedent conditions and events that may have influenced respondents' career decisions. The data can also be merged with results from the fifth follow-up questionnaire, which included special questions related to teaching.

The history of members of the Class of 1972 from their high school years through their early 30s is widely considered as the baseline against which the progress and achievements of subsequent cohorts are to be measured. Researchers have drawn on this archive since its inception. To date, the principal comparisons have been with the other two NELS studies: High School and Beyond (HS&B) and the National Education Longitudinal Study of 1988 (NELS:88). These three studies together provide a particularly rich resource for examining the changes that have occurred in American education during the past 20 years. Data from these studies can be used to examine how student academic coursework, achievement, values, and aspirations have changed, or remained constant, throughout this period.

The NELS studies offer a number of possible time points for comparison. Cohorts can be compared on an *intergenerational or cross-cohort time-lag basis.* Both cross-sectional and longitudinal time-lag comparisons are possible. For example, cross-sectionally, NLS-72 seniors in 1972 can be compared to HS&B base-year seniors in 1980 and to NELS:88 second follow-up seniors in 1992. Longitudinally, changes measured between the senior year and 2 years after graduation can be compared across studies. *Fixed time comparisons* are also possible; groups within each study can be compared to each other at different ages though at the same point in time. Thus, NLS-72 seniors, HS&B seniors, and HS&B sophomores can all be compared in 1986—some 14, 6, and 4 years after each respective cohort completed high school. Finally, *longitudinal comparative analyses* of the cohorts can be performed by modeling the history of the age/grade cohorts. The possible comparison points and the considerations of content and design which may affect the comparability of data across the cohorts are discussed in *National Education Longitudinal Study of 1988: Trends Among High School Seniors, 1972–1992* (NCES 95-380).

3. KEY CONCEPTS

A few key terms relating to NLS-72 are defined below.

Test Battery. Six cognitive tests administered during the base year: (1) Vocabulary (15 items, 5 minutes), a brief test using a synonym format; (2) Picture Number (30 items, 10 minutes), a test of associative memory consisting of a series of drawings of familiar objects, each paired with a number; (3) Reading (20 items, 15 minutes), a test of comprehension of short passages; (4) Letter Groups (25 items, 15 minutes), a test of inductive reasoning which required the student to draw general concepts from sets of data or to form and try out hypotheses in a nonverbal context; (5) Mathematics (25 items, 15 minutes), a quantitative comparison in which the student indicated which of two quantities was greater, or asserted their equality or the lack of sufficient data to determine which quantity was greater; and (6) Mosaic Comparisons (116 items, 9 minutes), a test measuring perceptual speed and accuracy through items which required detection of small differences between pairs of otherwise identical mosaics or tile-like patterns.

Socioeconomic Status (SES). A composite scale developed as a sum of standardized scales of father's education, mother's education, 1972 family income, father's occupation, and household items. The latter two underlying scales were computed from base-year Student Questionnaire responses. The other three underlying scales were derived from base-year responses as augmented by first follow-up responses and responses to a second follow-up resurvey to obtain this (and other) information from sample members who had failed to provide it previously. Each index component was first subjected to factor analysis that revealed a common factor with approximately equal weights for each component. Each of the components was then standardized, and an equally weighted combination of the five standard scores yielded the SES composite. The data file contains both the raw score and a categorized SES score (SES Index).

4. SURVEY DESIGN

Target Population

The population of students who, in spring 1972, were 12th graders (high school seniors) in public and private schools located in the 50 states and the District of Columbia. Excluded were students in schools for the physically or mentally handicapped, students in schools for legally confined students, early (mid-year) graduates, dropouts, and individuals attending adult education classes.

Sample Design

The NLS-72 sample was designed to be representative of the approximately 3 million high school seniors enrolled in more than 17,000 schools in the United States in spring 1972. The base-year sample design was a stratified, two-stage probability sample of students from all public and private schools, in the 50 states and the District of Columbia, which enrolled 12th graders during the 1971–1972 school year. Excluded were schools for the physically or mentally handicapped and schools for legally confined students. A sample of schools was selected in the first stage. In the second stage, a random sample of 18 high school seniors was selected within each participating school.

The base-year first-stage sampling frame was constructed from computerized school files maintained by the U.S. Department of Education and the National Catholic Educational Association. The original sampling frame called for 1,200 schools; that is, 600 strata with two schools per stratum. The strata were defined based upon the following variables: type of control (public or private), geographic region, grade 12 enrollment size, geographic proximity to institutions of higher education, proportion of minority group enrollment (for public schools only), income level of the community, and degree of urbanization. Schools were selected with equal probabilities for all but the smallest size stratum (schools with enrollment under 300). In that stratum, schools were selected with probability proportional to enrollment. All selections were without replacement. To produce sufficient sizes for intensive study of disadvantaged students, schools in low-income areas and schools with high proportions of minority group enrollment were sampled at twice the rate used for the remaining schools. Within each stratum, four schools were selected, and then two of the four were randomly designated as the primary selections. The other two schools were retained as backup or substitute selections (for use only if one or both of the primary schools did not cooperate).

The second stage of the base-year sampling procedure consisted of first drawing a simple random sample of 18 students per school (or all if fewer than 18 were available) and then selecting 5 additional students (if available) as possible replacements for nonparticipants. In both cases, the students within a school were sampled with equal probabilities and without replacement. Dropouts, early (mid-year) graduates, and those attending adult education classes were excluded from the sample. The oversampling of schools in low-income areas and schools with relatively high minority enrollment led to oversampling of low-income and minority students.

BEST COPY AVAILABLE

Sample redefinitions and augmentations. At the close of the base-year survey, 1,043 (948 primary schools and 95 backup schools) of a targeted 1,200 schools and an additional 26 "extra" backup schools had participated (school participation being defined as students from that school contributing SRIFs, Test Batteries, or Student Questionnaires). A backup school was termed "extra" if, ultimately, both primary sample schools from that stratum also participated. An additional 21 primary schools indicated that they had no 1972 seniors. At this point, there remained several strata with no participating schools and many more with only one school. To reduce the effects of the large base-year school nonresponse, a resurvey activity was implemented in the summer of 1973 prior to the first follow-up survey. An attempt was made to elicit cooperation from the 231 nonparticipating base-year primary schools and to obtain replacement schools to fill empty or partially filled strata utilizing backup schools if necessary. The resurvey was successful in 205 of the 231 primary sample schools. Students from 36 backup schools were also included so as to obtain at least two participating schools in the first follow-up survey from each of the 600 original strata. Students from the 26 "extra" base-year schools were not surveyed during the first follow up; however, 18 of the 26 "extra" schools were included in the second and subsequent follow-up surveys to avoid elimination of cases with complete base-year data.

To compensate for base-year school undercoverage, samples of former 1972 senior students were selected for inclusion in the first and subsequent follow ups from 16 sample augmentation schools (8 new strata); these schools were selected from those identified in 200 sample school districts canvassed to identify public schools not included in the original sampling frame. As before, 18 students per school were selected (as feasible) by simple random sample.

The number of students in the final sample from each sample school was taken as the number of students who were offered a chance to be in the sample and who also were eligible. This included all sample eligibles, both respondents and nonrespondents, but excluded students who were not eligible for the study—such as dropouts, early (mid-year) graduates, and those attending adult education classes. The final NLS-72 sample included 23,451 former 1972 seniors and 1,339 sample schools—1,153 participating primary schools, 21 primary schools with no 1972 seniors, 131 backup sample schools, 18 "extra" schools in which base-year student data had been completed, and 16 augmentation schools.

Retests of a subset of the base-year Test Battery were targeted for a subsample of 1,016 of the 14,628 eligible fourth follow-up sample members who had completed both a Student Questionnaire and a Test Battery in the base-year survey. Because a self-weighting subsample would have yielded an inadequate number of Black subsample members, a design option that oversampled Blacks was adopted. In addition to the stratification by race, the sample was controlled within strata on three factors believed to be highly correlated with retest ability scores: base-year ability, socioeconomic status, and postsecondary educational achievement. The control was achieved by applying an implicit stratification procedure. Test results were obtained from 692 of those in the subsample. Additional retest data were requested for all fourth follow-up sample members who had participated in the base-year testing and who were scheduled for a personal interview. This resulted in additional test data for 1,956 individuals (50.3 percent of those defined as request-eligible).

Fifth Follow-up Survey. The fifth follow-up sample was an unequal probability subsample of the 22,652 students who had participated in at least one of the five previous waves of NLS-72. The fifth follow up retained the essential features of the initial stratified multistage design but differed from the base-year design in that the secondary sampling unit selection probabilities were unequal, whereas they were equal in the base-year design. This inequality of selection probabilities allowed oversampling of policy-relevant groups and enabled favorable cost-efficiency tradeoffs.

In general, the retention probabilities for students were inversely proportional to the initial sample selection probabilities. The exceptions were: (1) sample members who were retained with certainty or at a higher rate than others because of their special policy relevance; (2) persons with very small initial selection probabilities who were retained with certainty; and (3) nonparticipants in the fourth follow up who were retained at a lower rate than other sample members because they were expected to be more expensive to locate and because they would be less useful for longitudinal analysis.

The subgroups of the original sample retained with certainty were: (1) Hispanics who participated in the fourth follow-up survey; (2) teachers and "potential teachers" who participated in the fourth follow-up survey (a "potential teacher" was defined as a person who majored in education in college or was certified to teach, or whose background was in the sciences); (3) persons with a

BEST COPY AVAILABLE

4-year or 5-year college degree or a more advanced degree; and (4) persons who were divorced, widowed, or separated from their spouses, or never-married parents. These groups overlapped and did not comprise distinct strata in the usual sense.

Teaching Supplement. The fifth follow-up sample included all sample members known to be teachers or potential teachers as of 1979 (the fourth follow up). To identify those sample members who had become teachers between the fourth and fifth follow ups, a direct question was included in the fifth follow-up main questionnaire. Respondents were selected for the Teaching Supplement sample if they indicated that they were (1) currently an elementary or secondary teacher, (2) formerly an elementary or secondary teacher, or (3) trained as an elementary or secondary teacher but never went into teaching. Of the 12,841 fifth follow-up respondents, 1,517 were eligible for the Teaching Supplement.

Postsecondary Education Transcript Study (PETS). In the first through fourth follow-up surveys, approximately 14,700 members of the NLS-72 cohort reported enrollment at one or more postsecondary institutions. An attempt was made to obtain a transcript from each school named by a respondent. Thus, no probabilistic sampling was done to define the PETS sample.

Data Collection and Processing

The base-year survey was administered through group administration. For the first four follow-up surveys, field operations began in the summer/fall of the survey year and continued through the spring of the following year; for example, the third follow-up survey (1976) data collection began in October 1976 and continued through June 1977. For the fifth follow-up survey, the data collection began in March 1986 and ended in mid-September 1986. The Educational Testing Service (ETS) administered the base-year survey; the Research Triangle Institute (RTI) carried out the first through fourth follow-up surveys; and the National Opinion Research Center (NORC) conducted the fifth follow-up survey.

Reference dates. Sample members in each of the first four follow-up surveys were asked about family information (marital status, spouse's status, number of children), location, and what they were doing with regard to work, education, and/or training during the first week of October of the survey year; fifth follow-up participants were asked the same questions for the first week of February 1986. Family income was requested for the preceding two years, and political and volunteer activities were requested for the past 24 months. Participants in each follow-up survey were also asked for summaries of educational and work experiences and activities for the intervening year(s) since the last survey. For the first four follow-up surveys, this information was requested as of the month of October in the intervening year(s) or sometimes overall for each year preceding the survey; fifth follow-up survey participants were asked detailed questions for up to four jobs and for attendance at up to two educational institutions since October 1979.

Data collection. Data collection instruments and procedures for the base-year survey were designed during the 1970–71 school year and were tested on a small sample of seniors in spring 1971. One year later, the full-scale NLS-72 study was initiated. Through an in-school group administration in the base year, each student was asked to complete a Test Battery measuring both verbal and nonverbal aptitude and to complete applicable portions of a Student Questionnaire containing 104 questions distributed over 11 major sections. Students were given the option of completing the Student Questionnaire in school or taking it home and answering the questions with the assistance of their parents. In addition, school administrators at each participating school were asked to complete a Student Record Information Form (SRIF) for each student in the sample and a School Questionnaire. One or two counselors from each school in the sample were asked to complete a Counselor Questionnaire.

Follow-up surveys. In fall 1973, 1974, 1976, and 1979 and spring 1986, sample members (or a subsample) were again contacted. After extensive tracing to update the name and address files, follow-up questionnaires were mailed to the last known addresses of sample members whose addresses appeared sufficient and correct and who had not been removed from active status by prior refusal, reported death, or other reason. Respondents to the third through fifth follow-ups were offered small monetary incentives for completing the questionnaires. These mailouts were followed by a planned sequence of reminder postcards, additional questionnaire mailings, reminder mailgrams (for the first four follow ups) and telephone calls, personal interviews, and, for the third to fifth follow ups only, telephone interviews to nonrespondents. During personal interviews, the entire questionnaire was administered. During telephone interviews conducted in the last three follow ups, only critical items that were suitable for telephone administration were administered. In order to make survey procedures comparable, respondents were asked to keep a copy of the questionnaire in front of them for both telephone and in-person interviews.

In all follow ups, returned questionnaire cases missing critical items were flagged during data entry, and data were retrieved by specially-trained telephone interviewers. Although most questions were of the forced-choice type, coding was required for the open-ended questions on occupation, industry, postsecondary school, field of study, state where marriage and divorce occurred, and relationship. Occupational and industry codes were obtained from the U.S. Department of Commerce, Bureau of the Census' *Classified Index of Industries and Occupations, 1970* and *Alphabetical Index of Industries and Occupations, 1970*. These same sources were used in all follow ups. Coding of the names of postsecondary schools attended by the respondents was accomplished by using codes taken from NCES' *Education Directory, Colleges and Universities*. Field of study information was coded using NCES' *A Classification of Instructional Programs* (CIP). In the fifth follow up, for the first time, all codes were loaded into a computer program for quicker access. Coders entered a given response, and the program displayed the corresponding numerical code.

Prior to the fifth follow up, all data were entered via direct access terminals. The fifth follow-up survey marked the first time that NLS-72 data were entered with a combination of keyed entry and optical scanning procedures. Using a computer-assisted data entry (CADE) system, operators were able to combine data entry with traditional editing procedures. All critical items and filter items (plus error-prone data like dollar amounts and numbers in general) were processed by CADE. The rest of the data were optically scanned.

Teaching Supplement. Data collection procedures used for the Teaching Supplement, administered concurrently with the fifth follow up, were similar to those used for the follow-up surveys.

Postsecondary Education Transcript Study (PETS). Packets of transcript survey materials were mailed to the postsecondary schools in July 1984, with a supplemental mailing in November 1984. Altogether, 24,431 transcripts were initially requested from 3,983 institutions for 14,759 NLS-72 sample members. Telephone follow up of nonresponding schools began in September 1984, when transcripts had been received from about two-thirds of the schools.

After investigating several alternatives, NORC adapted its CADE system for processing postsecondary transcripts. A single member of the specially-trained data preparation staff analyzed the transcript document to determine its general organization and special characteristics; abstracted standard information from the highly varied documents into a common format; assigned standard numerical codes to such transcript data elements as major and minor fields of study, degrees earned, types of academic term, titles of courses taken, grades and credits; and entered all pertinent information into a computer file. Combining these steps ensured that transcripts would be handled as internally consistent, integrated records of an individual's educational activity. Moreover, since all transcript processing occurred at a single station, the use of CADE reduced the number of steps at which records might be lost or misrouted, or other errors introduced into the database.

Editing. For the base-year through fourth follow-up surveys, an extensive manual or machine edit of all NLS-72 data was conducted in preparing the release file for public use. Editing involved rigorous consistency checking of all routing patterns within an instrument (not just skip patterns containing "key" or critical items), as well as range checks for all items and the assignment of error or missing data codes as necessary. Checks of the hardcopy sources were required in some cases for error resolution.

Unlike the earlier surveys, all editing for the fifth follow up was carried out as part of CADE. The machine-editing steps used in the prior follow ups were implemented for scanned items. Since most of the filter questions in the fifth follow up were CADE-designated items, there were few filter-dependent inconsistencies to be handled in machine editing. Validation procedures for the fifth follow up centered on verification of data quality through item checks and verification of the method of administration for 10 percent of each telephone or personal interviewer's work. Field managers telephoned the respondent to check several items of fact and to confirm that the interviewer had conducted a personal or telephone interview, or had picked up a questionnaire. No cases failed validation.

Postsecondary Education Transcript Study (PETS). The CADE program enforced predetermined range and value limitations on each field. The program performed three types of error-screening: (1) through a check-digit system, the program disallowed entry of incorrect identification data (school FICE codes, student ID numbers, and combinations of schools and students); (2) each data field was programmed to disallow entry of illogical or otherwise incorrect data; and (3) each CIP code selected to classify a field of study or a course was confirmed by automatically displaying the CIP program name for the code next to the name (from the original CADE transcript) that the coder had entered. A sample of CADE transcripts was selected and printed from every completed data disk for supervisory review.

Estimation Methods

Weighting was in NLS-72 to adjust for sampling and nonresponse. Various composite variables have also been computed to assist in data analyses.

Weighting. The weighting procedures used for the various NLS-72 survey data are described below.

Student files. NLS-72 student weights are based upon the inverse of the probabilities of selection through all stages of the sampling process and upon nonresponse adjustment factors computed within weighting classes. Unadjusted raw weights—the inverses of sample inclusion probabilities—were calculated for all students sampled in each survey year. These weights are a function of the school selection probabilities and the student selection probabilities within school. The raw weight for a case equals the raw weight for the base-year sample divided by the conditional probability of selection into that follow-up survey, given that the case was selected into the base-year sample.

Because of the various sample redefinitions and augmentations and nonresponse to the various student instruments, several sets of adjusted weights were computed for each NLS-72 survey wave. Each weight is appropriate for a particular respondent group. The general adjustment procedure used was a weighting class approach, which distributes the weights of nonrespondents to respondents who are in the same weighting class. The adjustment involves partitioning the entire student sample (respondents and nonrespondents) into weighting classes (homogeneous groups with respect to survey classification variables), and performing the adjustments within weighting class. Adjusted weights for nonrespondents are set to 0, and their adjusted weights are distributed to respondents proportionally to the respondents' unadjusted weights. Differential response rates for students in different weighting classes are reflected in the adjustment, and the weight total within each weighting class (and thus for the sample as a whole) is maintained.

The weighting class cells were defined by cross-classifying cases by several variables. For the first through fourth follow-up surveys, the weighting class cells were: sex, race, high school program, high school grade point average, and parents' education. For the fifth follow-up survey, the weighting class cells were similar except that postsecondary education attendance was substituted for parents' education. In some instances, cells were combined by pooling across certain weighting class cells.

The third and fourth follow-up adjusted weights are applicable only to key items of these questionnaires or specified combinations of those items with items from other instruments. The restriction is related to a change in data collection procedures. One or two item nonresponse adjustment factors were calculated for each of these two surveys for the nonkey items that were not asked on the telephone. The appropriate adjusted weight for these two surveys should be multiplied by its nonresponse adjustment factor to provide a new weight that is appropriate to items on that questionnaire that are not key or combinations of such nonkey items with items from other instruments.

Refer to the NLS-72 user's manuals for complete weighting procedures and a specification of available weights and appropriate variables to which the weights apply.

Teaching Supplement file. One set of weights was specifically developed to compensate for unequal probabilities of retention in the Teaching Supplement sample and to adjust for nonresponse. Theoretically, the weights project to the population of high school seniors of 1972 who have taught elementary or secondary school or who were trained to teach but never went into teaching. The weighting procedures were similar to those used in the follow-up surveys and consisted of two basic steps. The first step was the calculation of a preliminary weight based on the inverse of the cumulative probabilities of selection for the Teaching Supplement. The preliminary weight for the Teaching Supplement is the fifth follow-up adjusted weight. The second weight carried out the adjustment of this preliminary weight to compensate for unit nonresponse. Respondents were cross-classified into weighting cells by race, high school grades, and status as a teacher (current or former teacher, or never taught).

School file. During the sequential determination of final school sampling memberships (including augmentations), several school sampling weights were computed. The principal purpose of the various school weights was to serve as a basis for subsequent computation of student weights as applicable to one or more of the several student instruments. Only two of the eight weights are of direct use in analyzing School File or other school-level data. The School File sample weight is appropriate for analyzing school-level data that potentially could be supplied by all 1,318 schools. This includes the School Questionnaire data.

The adjusted counselor weight should be used only in analyzing the responses to the Counselor Questionnaire; however, care must be exercised when analyzing these data. This questionnaire was only administered at base-year responding schools, and data were collected from either one or two counselors at each school.

Postsecondary Education Transcript Study (PETS) file. Because the PETS did not introduce any additional subsampling into the NLS-72 sample design, it was not necessary to calculate a new raw weight for this study. Instead the raw weight for the base-year survey was used. Three adjusted weights were created specifically for the analysis of transcript data. They are not meant to be associated with individual transcripts, but rather with all data for a particular individual. The first weight is a simple adjustment for nonresponse to the transcript study itself, where response is defined as an eligible case having one or more coded transcript records in the data file. The other two adjusted weights account for multiple instances of nonresponse (e.g., no transcripts, no response to the fourth follow-up survey, missing data for critical items). Nonresponse adjustments were computed as ratio adjustments within 39 separate weighting classes. Cases were assigned to each weight class based on sex, race/ethnicity, high school grades, and high school program, and within each group by whether or not only proprietary school(s) were attended. The final adjusted weights are the product of the raw weight for the "completed" case and the nonresponse adjustment factor for the weighting class to which the case belongs.

Imputation. The problem of missing data was resolved for certain items by supplemental data collections, the creation of composite variables, and some imputation of activity state and other variables. Most of the variables were created by pooling information from various items. For example, the activity states for 1972 and 1973 were updated with information gleaned from the Activity State Questionnaires that were administered concurrently with second follow-up operations. While some procedures for imputing missing data for activity state variables were incorporated in the steps of defining and recoding variables, two further phases of imputation procedures were implemented. The first phase involved direct logical inferences (e.g., type of school from name and address of school); the second phase involved indirect logical inferences (e.g., impute studying full-time for those whose study time is unknown, but who are studying and not working).

5. DATA QUALITY AND COMPARABILITY

The survey was implemented after an extensive period of planning, which included the design and field test of survey instrumentation and procedures. Any additional questions were field-tested prior to inclusion in the survey. The NLS-72 sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. Quality control activities were used throughout the data collection and processing of the survey.

Sampling Error

Statistical estimates derived from the NLS-72 survey data are subject to sampling variability. Like almost all national samples, the NLS-72 sample is not a simple random sample. Taylor Series estimation techniques were used to compute standard errors in published NLS-72 reports.

It is often useful to report design effects and the root mean design effect in addition to standard errors for complex surveys such as NLS-72. Results from several NLS-72 studies suggest that a straightforward multiplicative adjustment of the simple random sample standard error equation adequately estimates the actual standard error estimate for a percentage. The three generalized mean design effects for the first, second, and third follow-up surveys are, respectively, the square root of 1.39, 1.35, and 1.44. To be conservative, the higher value—the square root of 1.44—can be used as an estimate for fourth follow-up data. For the NLS-72 fifth follow up, the mean design effect for the overall NLS-72 sample is 2.64. The mean design effects indicate that an estimated percentage in the NLS-72 data is—on average—more than twice as variable as the corresponding statistic from a simple random sample of the same size. The mean design effects vary across the domains from a low of 2.0 for the respondents from the highest socioeconomic (SES) quartile to a high of 3.8 for Black respondents.

Nonsampling Error

The major sources of nonsampling error in NLS-72 were coverage error and nonresponse error.

Coverage error. To identify public schools not included in the original sample frame, an additional sample of 200 school districts was contacted after the base-year survey was completed. Forty-five additional schools were identified. To compensate for the base-year undercoverage, samples of former 1972 senior students from 16 of these

BEST COPY AVAILABLE

"augmentation" schools were included in the first and subsequent follow-up surveys. In addition, at the end of the base-year survey, several strata had no participating schools and many more had only one school (out of two planned in the original sample design). To compensate for this large school nonresponse, 205 base-year noncooperating primary schools and 36 additional backup schools were added to the sample prior to the first follow-up survey for "resurveying" with the original design. The former 1972 seniors from these augmented and resurveyed schools were asked some retrospective (senior year) questions during the first follow-up survey. These individuals—who redress the school frame undercoverage bias in the base year—do not appear on the NLS-72 base-year files that would typically be employed for comparisons of high school seniors, although the presence of some retrospective data for these individuals permits refinement of comparisons grounded in 1972 data.

Also, while every effort was made to include in the fifth follow up all persons who experienced teaching, it is conceivable that some individuals who entered teaching late were among the 6,000 cases not included in the fifth follow-up subsample. These individuals would not have had a chance to participate in the Teaching Supplement.

Nonresponse error. Detailed rates of response to various surveys and the availability of specific data items are provided in NLS-72 user's manuals.

Unit nonresponse. For the NLS-72 student surveys, there were two stages of sample selection and hence two types of unit nonresponse—school and student. During the base year, sample schools were asked to permit selection of individual seniors from the schools for the collection of questionnaire and test data. Schools that refused to cooperate in either of these activities were dropped from the sample. The bias introduced by base-year school-level refusals is of particular concern since it carried over into successive rounds of the survey. To the extent that the students in refusal schools differed from students in cooperating schools during later survey waves, the bias introduced by base-year school nonresponse persisted from one wave to the next. (Base-year school nonresponse is addressed under "Coverage error" above.)

Also, individual students at cooperating schools could fail to take part in the base-year survey. Student nonresponse would not necessarily carry over into subsequent waves since student nonrespondents in the base year remained eligible for sampling throughout the study. However, a study of third follow-up responses indicated that response to earlier survey waves was the most important predictor of response to the third follow up.

Due to intensive data collection procedures, the response rates to the individual NLS-72 surveys were high (80 percent or better) among eligible sample members. At the conclusion of fourth follow-up activities, a total of 12,980 individuals had provided information on each of the first five questionnaires (base-year and all four follow-up surveys), representing 78 percent of the 16,683 base-year respondents. As a result of the various retrospective data collection efforts, the number of individuals with some key data elements for all time points through the fourth follow-up survey is 16,450—73 percent of the 22,652 respondents who participated in at least one survey. In conjunction with the supplemental data collection efforts, this led to a high degree of sample integrity among the key longitudinal data elements.

Only sample members who had participated in at least one of the previous five waves were eligible for selection into the fifth follow-up sample. Of the 14,431 fifth follow-up sample members (excluding the deceased), 89.0 percent (unweighted) completed questionnaires in the fifth follow up; 92.2 percent participated in at least five of the six waves; and 62.1 percent participated in all six waves. There was moderate variation in weighted nonresponse rates by region; nonresponse was greater in the West and Northeast regions, lower in the South, and lowest in the North Central region. The relationship between urbanization and nonresponse was about the same as region—13 percent for rural schools, 15 percent for urban schools, and 18 percent for suburban schools. There was marked variation in nonresponse by race; Blacks showed the highest nonresponse (22.1 percent), followed closely by Hispanics (19.8 percent) and Whites (14.0 percent). Males had a higher nonresponse rate (17.3 percent) than females (13.6 percent).

In PETS, one or more transcripts were received for 91.1 percent of the 13,831 sample members reporting postsecondary school attendance since leaving high school. A single transcript was received for 55 percent of this group, two transcripts for 27 percent, and three or more transcripts for over 9 percent. At the transcript level, 87 percent of the 21,866 "in-scope" transcripts requested were supplied by the postsecondary schools (2,565 of the 24,431 transcripts initially requested could not be obtained because the school had no record of the student's attendance). Response rates varied from a high of 93 percent for transcripts sought from public 4-year colleges and universities to a low of 55 percent from the voca-

BEST COPY AVAILABLE

tional and proprietary schools. The higher response rates for the public and private nonvocational schools may be attributable to their typically longer period of existence and the relative permanence of their student files. Telephone follow-up calls to nonresponding schools revealed that nearly half of the vocational school transcripts requested for NLS-72 students were unavailable.

Item nonresponse. While unit nonresponse can be adjusted for by weighting, this approach is impractical for item nonresponse. Researchers should take into account that NLS-72 respondents often skipped questions incorrectly or gave unrecognizable answers. However, efforts were made to retrieve missing data for critical items by telephone, with a success rate of over 90 percent.

Most item nonresponse in NLS-72 resulted from respondents' limited recall of past events or misinterpretation of questions and routing instructions. Many items in the Student Files appear to have high (greater than 10 percent) nonresponse. In most instances, these items are associated with the routing patterns in the instruments. (A routing question is one that implicitly or explicitly directs a respondent around other questions in the instrument, e.g., skip patterns.) Rather conservative rules were used to label blanks as either missing (illegitimate skip—code 98) or inapplicable (legitimate skip—code 99). With the more complex routing patterns, a large section of items was sometimes coded illegitimate (code 98) due to just one inconsistency in the pattern. The user should be careful in interpreting data coded 98 and 99. When analysis requires data that lie within complex routing patterns, it is advisable to further examine the data within the routing items. Similarly, data labeled as suspect during the editing stage should be reexamined and possibly reclassified for specific analytic purposes.

Measurement error. The survey data were monitored for quality of processing and evaluated to determine the extent of any problems and the sources of errors. Some examples are given below.

Study of edit failures. If the respondent failed to answer certain key items properly, the questionnaire failed an edit and the respondent was contacted by telephone. A special study of survey responses in the third follow up was conducted to determine why so many questionnaires (over 60 percent) failed the edit process. This study concluded that: (1) the majority of edit failures associated with itemized financial questions involved the respondent's failure to supply answers to each of the requested line items; (2) items structured as "check all responses that apply" were likely to be failed by a substantial number of respondents; and (3) overall data entry errors were low except for items requiring itemized financial information.

Review of routing patterns. Quality control, completeness, routing, and consistency indices were created for use with the Student Files. Routing indices, computed identically for each survey, indicate the percentage of the routing questions that were ambiguously answered by an individual for a given instrument. The first four follow-up questionnaires contained 33, 52, 67, and 61 routine patterns, respectively. In general, 56–68 percent of all respondents proceeded through an instrument without violating any routing patterns; about 20–30 percent violated 1–5 routing patterns; and 7–15 percent violated 6–10 patterns. In all four instruments, there was a small number (3–7 percent) of sample members who had great difficulty with the routing patterns and violated the routing instructions in more than 10 different patterns.

Monitoring of data entry. For the first through fourth follow-up surveys, direct data entry terminals were used to key the survey data. Data entry error rates were computed for the fourth follow-up survey based on three keyings. After the initial keying, a random sample of questionnaires from each batch was selected for rekeying by two additional operators. The results were within the overall error rate tolerance established for NLS-72. The variable error rate across samples and operators on the selected supplemental questionnaires was 0.00040; the estimated character error rate was 0.00023.

Data Comparability

One of the major goals of the NELS Program is to make the data sufficiently comparable to allow cross-cohort comparisons between studies (NLS-72 vs. HS&B vs. NELS:88), as well as comparative analyses of data across waves of the same study. Nevertheless, the user should be aware of some variations in sample design, questionnaire and test content, and data collection methods that could impact the drawing of valid comparisons.

Sample design changes. Although the general NLS-72 sample design was similar for all waves, there were some differences worth noting. The original sample design called for two schools to be surveyed from each of 600 strata; however, at the end of the base-year survey, several strata had no participants and many more had only one. As a result of a resurvey effort during the first follow-up survey, the final sample included *at least* two participating schools from each stratum. The fifth follow-up sample design differed from the base-year design in that the student selection probabilities were equal in the base-year design but unequal in the fifth follow up.

Reporting period differences. The first four follow ups requested data as of October of the survey year, whereas the fifth follow up used February 1986 as the reference date.

Content changes. Due to the increased interest in event history analysis, the fifth follow-up survey collected more detailed information than did earlier surveys on the time periods during which respondents held jobs or were in school. Instead of recording one start and stop date for each school and job, up to eight time periods (or start and stop dates) were shown. To allow for maximum user flexibility, the responses were coded into pairs of start and stop dates.

Comparisons between NLS-72 student data and PETS data. There are substantial discrepancies between student-reported postsecondary attendance in the NLS-72 follow-up surveys and the evidence obtained from official school transcripts collected in the Postsecondary Education Transcript Study. One interpretation is that NLS-72 respondents overreported instances of postsecondary school attendance by about 10 percent (unweighted). If so, researchers analyzing postsecondary schooling using only the survey data would overestimate significantly the extent of this activity. Coding errors could offer further explanation for the discrepancies.

Comparisons with HS&B and NELS:88. The three NELS studies—NLS-72, HS&B, and NELS:88—were specifically designed to facilitate comparisons with each other. At the student level, three different kinds of comparative analyses are possible. (See section 2, Uses of Data for more detail.) The overall sample design is similar and a core of questionnaire items is comparable across all three studies. Additionally, item response theory methods can be used to place mathematics, vocabulary, and reading scores on the same scale for 1972, 1980, and 1982 seniors.

However, despite the considerable similarity between the NLS-72, HS&B, and NELS:88 studies, the differences in sample definition and statistical design have implications for intercohort analysis. Also, sampling error tends to be a greater problem for intercohort comparisons than for intracohort comparisons because there is sampling error each time an independent sample is drawn. In addition, a number of nonsampling errors may arise when estimating trends based on results from two or more sample surveys. For example, student response rates differed across the three NELS studies, and the characteristics of the nonrespondents may have differed as well. The accuracy of intercohort comparisons may also be influenced by differences in context and question order for trend items in the various student questionnaires; differences in test format, content, and context; and other factors such as differences in data collection and methodology. While some effort was made to maintain trend items over time in the NELS studies, strict test and questionnaire overlap was not considerable across the three studies. More specifically, differences exist in questionnaire construction and in mode and type of survey administration. See chapter 8 (HS&B) and chapter 6 (NELS:88) for additional information on the comparability of the three NELS studies.

6. CONTACT INFORMATION

For content information on NLS-72, contact:

Aurora D'Amico
Phone: (202) 502–7334
E-mail: aurora.d'amico@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

National Education Longitudinal Study of 1988: Second Follow-up Research and Development Working Papers, NCES 94–251, by P. Quinn, Washington, DC: 1995.

National Longitudinal Study of the High School Class of 1972: "Fourth Follow-Up Survey" Final Methodological Report, ED 217–052, by J.A. Riccobono, ed., et al. Washington, DC: 1981.

National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-Up Survey" Data File User's Manual, CS 87–406c, by R. Tourangeau, P. Sebring, B. Campbell, M. Glusberg, B. Spencer, and M. Singleton. Washington, DC: 1987.

National Longitudinal Study of the High School Class of 1972: An Historical Overview and Summary, ED 217–051, by K. McAdams, et al. Washington, DC: 1981.

The National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-Up Survey and High School and Beyond Third Follow-Up Survey: Field Test Report, ED 269–465, by C. Jones, et al. Washington, DC: 1985.

Uses of Data

National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors, NCES Working Paper 95–05, by S. Ingels and J. Baldridge. Washington, DC: 1995.

Survey Design

The National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-up (1986) Sample Design Report, CS 88–403c, by B. Spencer, P. Sebring, and B. Campbell. Washington, DC: 1987.

Psychometric Analysis of the NLS-72 and the High School and Beyond Test Batteries, NCES 85–218, by D.A. Rock, T.L. Hilton, J.M. Pollack, R.B. Ekstrom, and M.E. Goertz. Washington, DC: 1985.

Sample Design for the Selection of a Sample of Schools with Twelfth-Graders for a Longitudinal Study, by Westat, Inc. Washington, DC: 1972.

Data Quality and Comparability

Bias Resulting from School Nonresponse: Methodology and Findings, by S.R. Williams and R.E. Folsom. Washington, DC: 1977.

Factors Associated with Edit Failure, NCES 82–213, by J.M. Wisenbaker. Washington, DC: 1981.

Factors Related to "Third Follow-Up" Survey Responses, NCES 82–209, by J.M. Wisenbaker and A.J. Kolstad. Washington, DC: 1981.

NLS Data Entry Quality Control: The "Fourth Follow-Up" Survey, ED 221–593, by L.B. Henderson and D.R. Allen. Washington, DC: 1981.

Chapter 8: High School and Beyond (HS&B) Longitudinal Study

1. OVERVIEW

The High School and Beyond (HS&B) Study was the second study conducted as part of NCES' National Longitudinal Studies Program. This program was established to study the educational, vocational, and personal development of young people, beginning with their elementary or high school years and following them over time as they take on adult roles and responsibilities. The HS&B Study included two high school cohorts—a senior cohort (the graduating class of 1980) and a sophomore cohort (the sophomore class of 1980). Students, school administrators, teachers, parents, and administrative records provided data for the study. HS&B results can be compared with the results of two other longitudinal studies—the National Longitudinal Study of the High School Class of 1972 (NLS-72) and the National Education Longitudinal Study of 1988 (NELS:88). (See chapters 7 and 6 for descriptions of these studies.)

The HS&B Study covered more than 30,000 high school seniors and 28,000 high school sophomores. It primarily consisted of a base year survey in 1980 and four follow-up surveys in 1982, 1984, 1986, and 1992. Record studies were also conducted to obtain key supplemental data on students. As part of the first follow up, high school transcripts were requested for the sophomore cohort, providing information on the sophomores' course-taking behavior through their 4 years of high school. Postsecondary transcripts were collected in 1984 for the senior cohort and in 1987 and 1993 for the sophomore cohort. In addition, student financial aid data were obtained from administrative records in 1984 for the senior cohort and in 1986 for the sophomore cohort. The HS&B project ended in 1993 after the completion of the fourth follow-up survey and related transcripts study of the sophomore cohort.

LONGITUDINAL SAMPLE SURVEY OF THE HIGH SCHOOL SOPHOMORE AND SENIOR CLASSES OF 1980; BASE-YEAR SURVEY AND FOUR FOLLOW UPS, ENDING IN 1992

HS&B collected data from:

- Students and dropouts
- School administrators
- Teachers
- Parents
- High school transcripts
- Postsecondary transcripts
- Postsecondary financial aid records

Purpose

To (1) study longitudinally *the given cohorts'* educational, vocational, and personal development, beginning with their high school years, and the personal, familial, social, institutional, and cultural factors that may affect that development; and (2) compare the results with data from the NLS-72 and NELS:88 studies to facilitate *cross-cohort* studies of American youth's schooling and socialization.

Components

The HS&B Study compiled data from a sample of students, parents, teachers, and school administrators in a base year and four follow-up surveys. It also collected high school and postsecondary transcripts and administrative financial aid records. The various components are described below.

Base Year Survey. The base year survey was conducted in spring 1980 and comprised the following:

Student Questionnaire. Students were asked to (1) fill out a Student Identification Pages booklet, which included several items on the use of non-English languages as well as confidential identifying information; (2) complete a questionnaire that focused on the student's individual and family background, high school experiences, work experiences, future educational plans, future occupational goals, and plans for and ability to finance postsecondary education; and (3) take timed cognitive tests that measured verbal and quantitative abilities. The sophomore test battery included achievement measures in science, writing, and civics, while seniors were asked to respond to tests measuring abstract and nonverbal abilities.

School Questionnaire. Completed by an official in the participating school, this questionnaire collected information about enrollment, staff, educational programs, facilities and services, dropout rates, and special programs for handicapped and disadvantaged students.

Teacher Comment Checklist. At each grade level, teachers had the opportunity to answer questions about the traits and behaviors of sampled students who had been in their classes. The typical student in the sample was rated by an average of four different teachers.

Parent Questionnaire. A sample of parents provided information about family attitudes, family income, employment, occupation, salary, financial planning, and how these affect postsecondary education and goals. The results include responses from the parents of about 3,600 sophomores and 3,600 seniors.

First Follow-up Survey. The first follow-up survey was conducted in spring 1982. As in the base survey, information was collected from students, school administrators, and parents. For the 1980 senior cohort, high school and postsecondary experiences were the main focus of the survey; seniors were asked about their school and employment experiences, family status, and attitudes and plans. For the 1980 sophomore cohort, the survey gathered information on school, family, work experiences, educational and occupational aspirations, personal values, and test scores of sample participants. A high school transcript collection was also part of the first follow up for sophomore cohort members. (See below for more detail.)

Sophomores were classified by high school status as of 1982 (i.e., dropout, same school, transfer, or early graduate). Dropouts completed a *Not Currently in High School Questionnaire,* which included some questions from the regular Student Questionnaire but focused on the student's reasons for dropping out and the impact on his/her educational and career development. In addition to the regular Student Questionnaire, a *Transfer Supplement* was completed by members of the sophomore cohort who had transferred out of the base year sample high school to another high school. This supplement gathered information on reasons for transferring and for selecting a particular school, length of interruption in schooling and reasons, and particulars about the school itself (type, location, entrance requirements, size of student body, grades). Sophomore cohort members who graduated from high school ahead of schedule completed an *Early Graduate Supplement* in addition to the regular questionnaire. The Early Graduate Supplement documented reasons for and circumstances of early graduation, adjustments required to finish early, and respondents' activities compared with those of other out-of-school survey members (i.e., dropouts, 1980 seniors).

Second Follow-up Survey. This survey was conducted in spring 1984. For both the sophomore and senior cohorts, the survey collected data on the student's work experience, postsecondary schooling, earnings, periods of unemployment, and so forth. For seniors, postsecondary transcripts and financial aid records were also collected. (See below for more detail.)

Third Follow-up Survey. This survey was administered in spring 1986, using the same questionnaire for both the sophomore and senior cohorts. To maintain comparability with prior waves, many questions from earlier follow-up surveys were repeated. Respondents were asked to update background information and to provide information about their work experience, unemployment history, education and other training, family information (including marriage patterns), income, and other experiences and opinions. Financial aid records and postsecondary transcripts were collected for sophomores. (See below for more detail.)

Fourth Follow-up Survey. This survey was administered in spring 1992 to only the sophomore cohort. The survey sought to obtain valuable information on issues of access to and choice of undergraduate and graduate educational institutions, persistence in obtaining educational goals, progress through the curriculum, rates of degree attainment and other assessments of educational outcomes, and rates of return to the individual and society. A second collection of postsecondary transcripts for sophomore cohort members took place in 1993. (See below for more detail.)

Record Studies. The following record studies were conducted during the course of the HS&B project.

High School Transcript Study. In fall 1982, as part of the first follow up, nearly 16,000 high school transcripts were collected for sophomore cohort students who were seniors in 1982. This data collection allows the study of the course-taking behavior of the sophomore cohort throughout their four years of high school. Data include a six-digit course number for each course taken; course credit, expressed in Carnegie units (a standard of measurement that represents one credit for the completion of a 1-year course); course grade; year course was taken; grade point average; days absent; and standardized test scores.

Postsecondary Education Transcript Study. This study gathered data on students' academic histories since leaving high school. As part of the second follow up in 1984, postsecondary transcripts were collected for the senior cohort. Transcripts were requested from all postsecondary institutions reported by senior cohort members in the first and second follow-up surveys. Transcript data include dates of attendance; fields of study; degrees earned; and the titles, grades, and credits of every course attempted at each institution.

In 1987 and again in 1993, postsecondary transcripts were collected for the sophomore cohort. The latter collection allowed information to be obtained on sophomore cohort members who had received their baccalaureate degrees and then went on to pursue graduate, doctoral, and first professional degrees.

Student Financial Aid Records. In 1984, HS&B collected institutional financial aid records and federal records of the Guaranteed and Student Loan Program and the Pell Grant Program for seniors who had indicated postsecondary attendance. The federal financial aid records were obtained for the sophomore cohort in 1986.

Periodicity

The base year survey was conducted in 1980, with four follow ups in 1982, 1984, 1986, and 1992 (only the sophomore cohort). High school transcripts were collected for the sophomore cohort in 1982. Postsecondary transcripts were collected for the senior cohort in 1984 and for the sophomore cohort in 1987 and 1993. Student financial aid records were collected for the senior cohort in 1984 and the sophomore cohort in 1986.

2. USES OF DATA

The HS&B Study provides information on the educational, vocational, and personal development of young people as they move from high school into postsecondary education or the workforce and then into adult life. The initial longitudinal study (NLS-72) laid the groundwork for comparison with HS&B. It recorded the economic and social conditions surrounding high school seniors in 1972 and, within that context, their hopes and plans; subsequently, it measured the outcomes while also observing the intervening processes. The HS&B base year survey of 1980 seniors is directly comparable to NLS-72 data on 1972 seniors. With the follow-up data, trend comparisons can be made for the period 1972 to 1984. (See *A Guide to Using NELS:88 Data*, by J. Owings et al.) By comparing the results of the HS&B and NLS-72 studies, researchers can determine how plans and outcomes differ in response to changing conditions, or remain the same despite such changes. HS&B permits researchers to further monitor change by, for example, measuring the economic returns of postsecondary education for minorities and delineating the need for financial aid.

The HS&B Study allows both cross-sectional and longitudinal analyses of the students who were sophomores or seniors in 1980. The data are used to address issues of educational attainment, employment, family formation, personal values, and community activities since 1980. For example, a major study on high school dropouts used HS&B data to demonstrate that a large number of dropouts return to school and earn a high school diploma or an equivalency certificate. Other examples of issues and questions that can be addressed are:

- How, when, and why do students enroll in postsecondary education institutions?
- Did those who (while in high school) expected to complete the baccalaureate degree actually do so?
- How has the percentage of recent graduates from a given cohort who enter the workforce in their field changed over the past years?
- What are the long-term effects of not completing high school in the traditional way? How do employment and earnings event histories of traditional high school graduates differ from those who did not finish high school in the traditional manner?
- Do individuals who attend college earn more than those who do not attend college? What is the effect of student financial aid?

- What percentage of college graduates is eligible or qualified to enter a public service profession such as teaching?
- How many enter the workforce full-time in the area for which they are qualified?
- How and in what ways do public and private schools differ?

3. KEY CONCEPTS

Some of the key terms related to HS&B are defined below.

Cognitive Tests. Achievement tests administered to both cohorts in the base year survey and to only sophomores in the first follow up. The content was as follows: (1) Vocabulary (21 items, 7 minutes), using a synonym format; (2) Reading (20 items, 15 minutes), consisting of short passages (100–200 words) followed by comprehension questions and a few analysis and interpretation items; (3) Mathematics (38 items, 21 minutes), in which students were asked to determine which of two quantities was greater, whether they were equal, or whether there was insufficient data to answer the question; (4) Science (20 items, 10 minutes), based on science knowledge and scientific reasoning ability; (5) Writing (17 items, 10 minutes), based on writing ability and knowledge of basic grammar; and (6) Civics Education (16 questions, 5 minutes), based on various principles of law, government, and social behavior.

Course Offering and Course Taking. Course-offering data were collected from the School Questionnaires filled out by school administrators; course offerings include regular and advanced placement curricula provided by the schools. Course-taking data were collected in different ways for the sophomore and senior cohorts. For sophomores, official high school transcripts provided records of students' coursework. For the senior cohort, high school transcripts were not available; instead, coursework was self-reported by seniors in a series of items asking retrospectively about the courses and hours taken. Despite these differences in data collection, the listings of courses for the two cohorts were consistent, including major subjects in both regular and advanced placement curricula.

Socioeconomic Status (SES). Indicated by a set of composite variables, constructed from base year and first follow-up data—using father's occupation, father's education, mother's education, family income, and material possessions in the household.

4. SURVEY DESIGN

Target Population

High school students who were in the 10th or 12th grade in U.S. public and private schools in spring 1980.

Sample Design

HS&B was designed to provide nationally representative data on 10th- and 12th-grade students in the United States.

Base Year Survey. In the base year, students were selected using a two-stage, stratified probability sample design, with secondary schools as the first-stage units and students within schools as the second-stage units. Sampling rates for each stratum were set so as to select in each stratum the number of schools needed to satisfy study design criteria regarding minimum sample sizes for certain types of schools. The following types of schools were oversampled to make the study more useful for policy analyses: public schools with a high percentage of Hispanic students; Catholic schools with a high percentage of minority group students; alternative public schools; and private schools with high achieving students. Thus, some schools had a high probability of inclusion in the sample (in some cases, equal to 1.0), while others had a low probability of inclusion. The total number of schools in the sample was 1,122, selected from a frame of 24,725 schools with grades 10 or 12 or both. Within each stratum, schools were selected with probabilities proportional to the estimated enrollment in their 10th and 12th grades.

Within each school, 36 seniors and 36 sophomores were randomly selected. In those schools with fewer than 36 seniors or 36 sophomores, all eligible students were drawn in the sample. Students in all but the special strata were selected with approximately equal probabilities. The students in special strata were selected with higher probabilities. Special efforts were made to identify sampled students who were twins or triplets so that their co-twins or co-triplets could be invited to participate in the study.

Substitution was carried out for schools that refused to participate in the survey. There was no substitution for students who refused, for students whose parents refused, or for students who were absent on Survey Day and makeup days.

First Follow-up Survey. The first follow-up sophomore and senior cohort samples were based on the base year samples, retaining the essential features of a stratified multistage design. (For details beyond those given below, see *High School and Beyond First Follow-Up (1982) Sample Design Report,* by R.E. Tourangeau, et al.)

For the sophomore cohort, all of the 1,015 schools selected for the base year sample were included in the first follow up except 40 schools that had no 1980 sophomores, had closed, or had merged with other schools in the sample. The sample also included 17 schools that received two or more students from base year schools; school-level data from these institutions were eventually added to students' records as contextual information. However, these schools were not added to the existing probability sample of schools.

The sophomores still enrolled in their original base year schools were retained with certainty since the base year clustered design made it relatively inexpensive to resurvey and retest them. Sophomores no longer attending their original base year schools were subsampled (i.e., dropouts, early graduates, students who transferred as individuals to a new school). Certain groups were retained with higher probabilities in order to support statistical research on such policy issues as excellence of education throughout the society, access to postsecondary education, and transition from school to the labor force.

Students who transferred as a class to a different school were considered to be still enrolled if their original school had been a junior high school, had closed, or had merged with another school. Students who had graduated early or had transferred as individuals to other schools were treated as school leavers for the purposes of sampling. The 1980 sophomore cohort school leavers were selected with certainty or according to predesignated rates designed to produce approximately the number of completed cases needed for each of several different sample categories. School leavers who did not participate in the base year were given a selection probability of 0.1.

For the 1980 senior cohort, students selected for the base year sample had a known, nonzero chance of being selected for the first and all subsequent follow-up surveys. The first follow-up sample consisted of 11,995 selections from the base year probability sample. This total included 11,500 selections from among the 28,240 base year participants and 495 selections from among the 6,741 base year nonparticipants. In addition, 204 nonsampled cotwins or co-triplets (who were not part of the probability sample) were included in the first follow-up sample, resulting in a total of 12,199 selections.

High School Transcript Study (1980 Sophomore Cohort). Subsequent to the first follow-up survey, high school transcripts were sought for a probability subsample of nearly 18,500 members of the 1980 sophomore cohort. The subsampling plan for the transcript study emphasized the retention of members of subgroups of special relevance for education policy analysis. Compared to the base year and first follow-up surveys, the transcript study sample design further increased the overrepresentation of racial and ethnic minorities, students who attended private high schools, school dropouts, transfers, early graduates, and students whose parents completed the base year Parent Questionnaire on financing postsecondary education. Transcripts were collected and processed for nearly 16,000 members of the sophomore cohort.

Second and Third Follow-up Surveys. The sample for the second follow-up survey of the 1980 sophomore cohort was based upon the design of the High School Transcript Study. A total of 14,825 cases were selected from among the nearly 18,500 retained for the transcript study. The second follow-up sample included disproportionate numbers of sample members from policy-relevant subpopulations. The members of the senior cohort selected into the second follow-up sample consisted exactly of those selected into the first follow-up sample. The senior and sophomore cohort samples for the third follow-up survey were the same as those used for the second follow up. The third follow up was the last survey conducted for the senior cohort. Postsecondary school transcripts were collected for all members of the senior cohort members who reported attending any form of postsecondary schooling in either of the follow-up surveys. Over 7,000 individuals reported more than 11,000 instances of postsecondary school attendance.

Fourth Follow-up Survey. The fourth follow up was composed solely of members from the sophomore cohort, and consisted exactly of those selected into the second and third follow-up sample. For any student who ever enrolled in postsecondary education, complete transcript information was requested from the institutions indicated by the student.

Data Collection and Processing

HS&B compiled data from six primary sources: students, school administrators, teachers, parents of selected students, high school administrative records (transcripts), and postsecondary administrative records (transcripts and financial aid). Data collection began in fall 1979 (when information from school administrators and teachers was first gathered) and ended in 1993 (when postsecondary transcripts of sophomore cohort members were collected). The National Opinion Research Center (NORC) at the University of Chicago was the contractor for the HS&B project.

Reference dates. In the base year survey, most questions referred to the student's experience up to the time of administration in spring 1980 (i.e., all 4 high school years for the senior cohort and the first 2 high school years for the sophomore cohort). In the follow ups, most questions referred to experiences that occurred between the previous survey and the current survey. For example, the second follow up largely covered the period between 1982 (when the first follow up was conducted) and 1984 (when the second follow up was conducted).

Data collection. In both the base year and first follow-up surveys, it was necessary to secure a commitment to participate in the study from the administrator of each sampled school. For public schools, the process began by contacting the chief state school officer. Once approval was gained at the state level, contact was made with District Superintendents and then with school principals. Wherever private schools were organized into an administrative hierarchy (e.g., Catholic school dioceses), approval was obtained at the superior level before approaching the school principal or headmaster. The principal of each cooperating school designated a School Coordinator to serve as a liaison between the NORC staff, school administrator, and selected students. The School Coordinator (most often a senior guidance counselor) handled all requests for data and materials, as well as all logistical arrangements for student-level data collection on the school premises.

In the 1980 base year survey, a single data collection method—on-campus administration—was used for both the sophomore and senior cohorts. In the first follow up, members of the sophomore cohort (nearly all of whom were then in the 12th grade) were resurveyed using methods similar to those of the base year survey. Since some of the 1980 sophomores had left school [illegible] 1982, the first follow-up survey involved on-campus administration for in-school respondents and off-campus group administration for school leavers (transfers, dropouts, early graduates). On-campus surveys generally were similar to those used in the base year. Off-campus survey sessions were held afterwards for school leavers in the sophomore cohort. Personal or telephone interviews were conducted with individuals who did not attend the sessions. Members of the 1980 senior cohort were surveyed primarily by mail. Nonrespondents to the mail survey (approximately 25 percent) were interviewed either in person or by telephone.

By the time of the second follow up, the sophomore cohort was out of school. In the second (1984) and third (1986) follow ups, data for both the sophomore and senior cohorts were collected through mailed questionnaires. Telephone and personal interviews were conducted with sample members who did not respond to the mailed survey within 2–3 months. Only the sophomore cohort was surveyed in the fourth follow up (1992). Computer-assisted telephone interviewing (CATI) was used to collect these data. The CATI program included two instruments; the first was used to locate and verify the identity of the respondent, while the second contained all of the survey questions. The average administration time for an interview was 30.6 minutes. Intensive telephone locating and field intervention procedures were used to locate respondents and conduct interviews.

Processing. Although procedures varied across survey waves, all Student Questionnaires in all waves were checked for missing critical items. Approximately 40 items in each of the main survey instruments were designated as critical or "key" items. Cases failed this edit if a codable response was missing for any of the key items. Such cases were flagged and then routed to the data retrieval station, where staff called respondents to obtain missing information or otherwise resolve the edit failure.

The base year procedures for data control and preparation differed significantly from those in the follow-up surveys. Since the base year student instruments were less complex than later instruments, the completed documents were sent directly from the schools to NORC's optical scanning subcontractor for conversion to machine-readable form. The scanning computer was programmed to perform the critical item edit on Student Questionnaires and to generate listings of cases missing critical data, which were then sent to NORC for data retrieval. School and Parent Questionnaires were converted to machine-readable form by the conventional key-to-disk method at NORC.

All follow-up questionnaires were sent to NORC for receipt control and data preparation prior to being shipped to the scanning subcontractor. The second follow-up survey contained optically scannable grids for the answers to numeric questions; staff examined numeric responses for correct entry (e.g., right justification, omission of decimal points). In the third follow up, a portion of the instrument was designed for computer-assisted data entry (CADE), while the rest was prepared for optical scanning. All major skip items and all critical items were entered by CADE. With this system, operators were able to combine data entry with the traditional editing procedures. The CADE system stepped question-by-question through critical and numeric items, skipping over questions that were slated for scanning and questions that were legitimately skipped because of a response to a filter question. Ranges were set for each question, preventing the accidental entry of illegitimate responses. CADE operators were also responsible for the critical item edit; those critical items that did not pass the edit were flagged for retrieval, both manually and by the CADE system. After the retrieved data were keyed, questionnaires were shipped to the scanning firm.

For the fourth follow up, a CATI system captured the data at the time of the interview. The CATI program examined the responses to completed questions and used that information to route the interviewer to the next appropriate question. It also applied the customary edits, described below under "Editing." At the conclusion of an interview, the completed case was deposited in the database ready for analysis. There was minimal post-data entry cleaning because the interviewing module itself conducted the majority of necessary edit checking and conversion functions. A CADE program was designed to enter and code transcript data.

The first through fourth follow ups required coding of open-ended responses on occupation and industry; postsecondary schools; major field of study for each postsecondary school; licenses, certificates, and other diplomas received; and military specialized schools, specialty, and pay grade. Coding was compatible with the coding done in NLS-72, using the same sources from NCES and the U.S. Bureau of the Census. (See chapter 7.) In the first follow up, staff also coded open-ended questions in the Early Graduate and Transfer Supplements, and transformed numeric responses to darkened ovals to facilitate optical scanning. In the third follow up, all codes were loaded into a computer program for more efficient access. Coders typed in a given response, and the program displayed the corresponding numeric code.

In the fourth follow up, interviewers received additional coding capabilities by temporarily exiting the CATI program and executing separate programs that assisted them in coding the open-ended responses. Data from the coding programs were automatically sent to the CATI program for inclusion in the data set. In addition to the online coding tasks, interviewers recorded verbatim descriptions of industry and occupation. The coding scheme for industry in the fourth follow up was a simplified version of the scheme used in previous rounds of HS&B (verbatims are available for more detailed coding). The coding scheme for occupation coding was adapted from verbatim responses received in the third follow up. Postsecondary institutions were coded with Federal Interagency Committee on Education (FICE) codes.

Editing. In addition to the critical item edit described above, a series of edits checked the data for out-of-range values and inconsistencies between related items. In the base year, machine editing was limited to examining responses for out-of-range values. No interim consistency checks were performed since there was only one skip pattern.

In the first and second follow ups, several sections of the questionnaire required respondents to follow skip instructions. Computer edits were performed to resolve inconsistencies between filter and dependent questions, detect illegal codes, and generate reports on the incidence of correctly and incorrectly answered questions. After improperly answered questions were converted to blanks, the student data were passed to another program for conversion to appropriate missing-data codes (e.g., "legitimate skip," "refused"). Detection of out-of-range codes was completed during scanning for all questions except those permitting an open-ended response. Hand-coded data for open-ended questions (occupation, industry, institution, field of study) were matched by computer against lists of valid codes.

In the third follow up, CADE carried out many of the steps that normally occur during machine editing. The system enforced skip patterns, range checking, and appropriate use of reserved codes—allowing operators to deal with problems or inconsistencies while they had the document in hand. For scanned items, the same machine-editing steps as those used in prior follow ups were implemented. Since most of the filter questions were CADE-designated items, there were few filter-dependent inconsistencies to be handled in machine editing.

In the fourth follow up, machine editing was replaced by the interactive edit capabilities of the CATI system, which tested responses for valid ranges, data field size, data type (numeric or text), and consistency with other answers or data from previous rounds. If the system detected an inconsistency due to a miskey by the interviewer, or if the respondent simply realized that he or she made a reporting error earlier in the interview, the interviewer could go back and change the earlier response. As the new response was entered, all of the edit checks performed at the first response were again performed. The system then worked its way forward through the questionnaire using the new value in all skip instructions, consistency checks, and the like until it reached the first unanswered question, and control was then returned to the interviewer. When problems were encountered, the system could suggest prompts for the interviewer to use in eliciting a better or more complete answer.

Estimation Methods

Weighting is used to adjust for sampling and unit nonresponse.

Weighting. The weights are based on the inverse of the selection probabilities at each stage of the sample selection process and on nonresponse adjustment factors computed within weighting cells. While each wave provided weights for statistical estimation, the fourth follow-up weights can illustrate the concept of weighting. The fourth follow up generated survey data and postsecondary transcript data. Weights were computed to account for nonresponse in both of these data collections.

First, a raw weight, unadjusted for nonresponse in any of the surveys, was calculated and included on the data file. The raw weight provides the basis for analysts to construct additional weights adjusted for the presence of any combination of data elements. However, ***caution should be used if the combination of data elements results in a sample with a high proportion of missing cases.*** For the survey data, two weights were computed. The first weight (was computed for all fourth follow-up respondents. The second weight was computed for all fourth follow-up respondents who also participated in the base year and first, second, and third follow-up surveys.

Two additional weights were computed to facilitate the use of the postsecondary transcript data. The collection of transcripts was based upon sophomore cohort reports of postsecondary attendance during either the third or fourth follow up. A student may have reported attendance at more than one school. The first transcript weight was computed for students for whom at least one transcript was obtained. It is therefore possible for a student who was not a respondent in the fourth follow up but who was a respondent in the third follow up, to have a nonzero value for the first transcript weight. The second transcript weight is more restrictive. It was designed to assign weights only to cases that were deemed to have complete data. Only students who responded during the fourth follow up (and hence students for whom a complete report of postsecondary education attendance was available and for whom all requested transcripts were received) were assigned a nonzero value for the second transcript weight. For students who did not complete the fourth follow-up interview, complete transcripts may have been obtained in the 1987 transcript study, but since it was not certain that these transcripts were complete, they were given a weight of zero.

Imputation. No imputation was performed in the HS&B Study.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Because the sample design for the HS&B cohorts involved stratification, disproportionate sampling of certain strata, and clustered probability sampling, the calculation of exact standard errors (an indication of sampling error) for survey estimates can be difficult and expensive.

Sampling error estimates for the first and second HS&B follow ups were calculated by the method of Balanced Repeated Replication (BRR) using BRRVAR, a Department of Education statistical subroutine. The BRR programs, WesVar and SUREG, are now available commercially. For the base year and the third and fourth follow ups, Taylor Series approximations were employed. More detailed discussions of the BRR and Taylor Series procedures can be found in the ***High School and Beyond Third Follow-Up Sample Design Report***, CS 88-402. The Data Analysis System (DAS), included as part of the public release file, automatically reports design-corrected Taylor Series standard errors for the tables it generates. Therefore, users of the DAS need make no adjustments to these estimates.

While design effects cannot be calculated for every estimate of interest to users, design effects will be similar from item to item within the same subgroup or population. Users can calculate approximate standard error estimates for items by multiplying the standard error under the simple random sample assumption by the square root of the average design effect for the population being studied.

Nonsampling Error

Nonsampling errors include coverage, nonresponse, and measurement errors.

Coverage error. Bias caused by explicit exclusion of certain groups of schools and students (e.g., special types of schools or students with disabilities or language barriers) is not addressed in HS&B technical reports. Potential coverage error in HS&B may relate to the exclusion of schools that refused to cooperate in the base year survey. Students who refused to participate in the base year survey were not excluded in the follow ups. Since students were randomly selected from the sampled schools, the HS&B sample design did not entail exclusion of specified groups. (See section 4, Sample Design.)

Nonresponse error.

Unit nonresponse. HS&B base year student-level estimates include two components of unit nonresponse bias: bias introduced by nonresponse at the school level, and bias introduced by nonresponse on the part of students attending cooperating schools. At the school level, some schools refused to participate in the base year survey. Substitution was carried out for refusal schools within stratum when there were two or more schools within the stratum. The bias introduced by base year school-level refusals is of particular concern since it carried over into successive rounds of the survey. Students attending refusal schools were not sampled during the base year and had no chance for selection into subsequent rounds of observation. To the extent that these students differed from students from cooperating schools in later waves of the study, the bias introduced by base year school nonresponse would persist. Student nonresponse did not carry over in this way since student nonrespondents remained eligible for sampling in later waves of the study.

In general, the lack of survey data for nonrespondents prevents the estimation of unit nonresponse bias. However, during the first follow up, School Questionnaire data were obtained from most of the base year refusal schools, and student data were obtained from most of the base year student nonrespondents selected for the first follow-up sample. These data provide a basis for assessing the magnitude of unit nonresponse bias in base year estimates.

Overall, 1,122 schools were selected in the original sample, and 811 of those schools (72 percent) participated in the survey. An additional 204 schools were drawn in a replacement sample. Student refusals and absences resulted in a weighted student completion rate of 88 percent in the base year survey. Participation was higher in most follow-up surveys. Completion rates in the first follow up were: 94 percent for seniors; 96 percent for sophomores eligible for on-campus survey administration; and 89 percent for sophomores who had left school between the base year and first follow up surveys (dropouts, transfer students, and early graduates). In the second follow up, 91 percent of senior cohort members and 92 percent of sophomore cohort members completed the survey. In the third follow up, completion rates were 88 percent for seniors and 91 percent for sophomores. Only the sophomore cohort was surveyed in the fourth follow up; 86 percent of the sample members participated.

As results from the fourth follow up illustrate, student nonresponse varied by demographic and educational characteristics. Males had a slightly higher nonresponse rate than females (a difference slightly over 3 percent). Blacks and Hispanics showed similarly high rates of nonresponse (around 20 percent), whereas nonresponse among White students was about 10 percent. Nonresponse increased as socioeconomic status decreased. Students who were in general or vocational programs during the base year were more likely to be nonrespondents than students in academic programs. Dropouts had higher nonresponse rates than other students. Students with lower grades and lower test scores showed higher nonresponse than students with higher grades and test scores. Students who were frequently absent from school showed higher nonresponse than students absent infrequently. Students with no postsecondary education by the time of the second follow up had higher nonresponse than students with some postsecondary education. By selected school characteristics, the highest nonresponse rates were among students from alternative public schools, schools with large enrollments, schools in urban areas, and schools in the Northeast and West.

The patterns were similar in earlier rounds of HS&B. Nonresponse analyses conducted by NORC support the following general conclusions:

(1) The school-level bias component in HS&B estimates is small, averaging less than 2 percent for base year and first follow-up estimates. It is probably of a similar magnitude for fourth follow-up estimates.

(2) The student-level bias component in base year estimates is also small, averaging about 0.5 percent for percentage estimates.

(3) The student-level bias component in first, second, and third follow-up estimates is limited by the nonresponse rates, which were about three-fourths of the base year rates.

(4) The student-level bias component in the fourth follow up is limited by the nonresponse rate, which was slightly higher than the base year rate.

The first and second conclusion together suggests that nonresponse bias is not a major contributor to error in base year estimates. The first and third suggest that nonresponse bias is not a major contributor to error in the first, second, and third follow-up estimates either. The first and fourth conclusion suggest that the fourth follow-up nonresponse bias might be a little greater than for the previous follow ups, but probably not by much. Each of these conclusions must be given some qualifications. The analysis of school-level nonresponse is based on data concerning the schools, not the students attending them. The analyses of student nonresponse are based on survey data and are themselves subject to nonresponse bias. Despite these limitations, the results consistently indicate that nonresponse had a small impact on base year and follow-up estimates.

Item nonresponse. Among students who participated in the survey, some did not complete the questionnaire or gave invalid responses to certain questions. The amount of item nonresponse varied considerably by item. For example, in the second follow up, a very low nonresponse rate of 0.1 percent was observed for a question asking whether the respondent had attended a postsecondary institution. A much higher nonresponse rate of 12.2 percent was obtained for a question asking if the respondent had used a micro or minicomputer in high school. Typical item nonresponse rates ranged from 3 to 4 percent.

Imputation was not used to compensate for item nonresponse in HS&B. However, an attempt was made in the fourth follow up to reduce item nonresponse. In previous rounds, interviews were conducted by self-administered questionnaires (SAQs). Unfortunately, respondents often skipped questions incorrectly or gave unrecognizable answers. Thus, more data were missing than would have occurred through personal interviewing. In the fourth follow up, interviewing was conducted using computer-assisted telephone interviewing (CATI). Unlike SAQs, CATI interviewing virtually eliminated missing data attributable to improperly skipped questions.

To evaluate the effectiveness of CATI interviewing, 25 items from both the third and fourth follow-up data were selected for comparison. Refusal and "don't know" responses were considered to be missing, but legitimate skips were not. For these 25 items, the overall percentage of missing items dropped from 4.36 percent in the third follow up to 1.88 percent in the fourth follow up. CATI also eliminated all multiple responses and resulted in uncodable verbatims for only the two income variables. In addition, more was known about the missing data in the fourth follow up. In the third follow up, only 7.2 percent of the missing data were classified as refusals or "don't know" responses. In the fourth follow up, 50.9 percent of the missing data were classified as refusals or "don't know" responses. The fact that most of the 25 comparisons showed a "very significant" decline in missing data supports a contention that missing data were reduced in the fourth follow up.

Measurement error. An examination of consistency between responses to the third and fourth follow ups provides an indication of the reliability of HS&B data.

Race/ethnicity. Race/ethnicity is one characteristic of the respondent that should not change between surveys. Overall, of the 12,309 respondents who reported their race/ethnicity on both questionnaires, 93.8 percent gave the same response in both years. However, certain race/ethnicity categories (e.g., Native American) had substantially less agreement. Only 53.4 percent of the respondents who classified themselves as Native Americans during the third follow up classified themselves as Native Americans again during the fourth follow up.

One explanation for these discrepancies may be the change in the method of survey administration. Unlike the third follow up, which involved self-administered questionnaires, the fourth follow up was conducted by telephone. The questionnaires mailed during the third follow up had the five race/ethnicity categories listed for the respondent to see. In the fourth follow up, respondents were simply asked over the telephone, "What is your race/ethnicity?" The interviewer coded the response. It is possible that Native Americans, Hispanics, and Asian/Pacific Islanders classified themselves as Black or White (not knowing that there was a more specific category for them), hence resulting in more Blacks and Whites in the fourth follow-up results.

Marital status. In the third follow up, respondents were asked about their marital status ***in the first week*** of February 1986. In the fourth follow up, respondents were asked about their marital status ***during and since*** February 1986. Although both questions asked about marital status during February 1986, respondents who had a change in marital status during the last three weeks of February could have given a different answer in the fourth follow up than in the third follow up. Overall, of the 11,854 respondents who gave their marital status on both questionnaires, 95.4 percent had answers that agreed.

Unlike the race/ethnicity question, memory and timing play an important role in matching answers for marital status. In this case, the recall period for third follow-up respondents was years shorter than the recall period for respondents in the fourth follow up. Respondents in the third follow up, which took place in spring 1986, were asked about a recent event. Respondents in the fourth follow up, which was conducted in spring 1992, were asked to recall their status back in February 1986. As with the race/ethnicity question, the method of administering the question differed between rounds—namely, the question formatting had changed and the fourth follow up used preloaded data to verify marital status.

Data Comparability

A goal of the National Longitudinal Studies Program is to allow comparative analysis of data generated in several waves of the same study and also to enable cross-cohort comparisons with the other longitudinal studies. While the HS&B and NLS-72 studies are largely compatible, a number of variations in sample design, questionnaires, and data collection methods should be noted to caution data users.

Comparability within HS&B. While many data items were highly compatible across waves, the focus of the questionnaires necessarily shifted over the years in response to the changes in the cohorts' life cycle and the concerns of education policymakers. For seniors in the base year survey and for sophomores in both the base year and first follow-up surveys, the emphasis was on secondary schooling. In subsequent follow ups, increasingly more items were collected dealing with postsecondary education and employment. Also, a major change in the data collection method occurred in the fourth follow up, when CATI was introduced as the primary approach. Earlier waves used mailed questionnaires supplemented by telephone and personal interviews.

Comparability with NLS-72. The HS&B Study was designed to build on NLS-72 in three ways. First, the HS&B base year survey included a 1980 cohort of high school seniors that was directly comparable to the NLS-72 cohort (1972 seniors). Replication of selected 1972 Student Questionnaire items and test items made it possible to analyze changes subsequent to 1972 and their relationship to federal education policies and programs in that period. Second, the introduction of the sophomore cohort in HS&B provided data on the many critical educational and vocational choices made between the sophomore and senior years in high school, thus permitting a fuller understanding of the secondary school experience and how it affects students. Third, HS&B expanded the NLS-72 focus by collecting data on a range of life cycle factors, such as family formation, labor force behavior, intellectual development, and social participation.

The sample design was largely similar for both the HS&B and NLS-72 studies, except that HS&B included a sophomore sample in addition to a senior sample. The questionnaires for the two studies contained a large number of identical or similar items dealing with secondary education and postsecondary work experience and education. The academic tests were also highly compatible. Of the 194 test items administered to the HS&B senior cohort in the base year, 86 percent were identical to items that had been given to NLS-72 base year respondents. Item response theory (IRT) was used in both studies to put math, vocabulary, and reading test scores on the same scale for 1972, 1980, and 1982 seniors. With the exception of CATI in the HS&B fourth follow up, both NLS-72 and HS&B used group administration of questionnaires and tests in the earliest surveys and mailed questionnaires in the follow ups. HS&B, however, involved more extensive efforts to supplement the mailings by telephone and personal interviews.

6. CONTACT INFORMATION

For content information on HS&B, contact:

Aurora M. D'Amico
Phone: (202) 502–7334
E-mail: aurora.d'amico@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

High School and Beyond Fourth Follow-Up Methodology Report, NCES 95–426, by D. Zahs, S. Pedlow, M. Morrissey, P. Marnell, and B. Nichols. Washington, DC: 1995.

Uses of Data

National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data, NCES Working Paper 95–06, by S. Ingels and J. Taylor. Washington, DC: 1995.

National Education Longitudinal Study of 1988: Conducting Trend Analyses: HS&B and NELS:88 Sophomore Cohort Dropouts, NCES Working Paper 95–07, by S. Ingels and K. Dowd. Washington, DC: 1995.

National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors, NCES Working Paper 95–05, by S. Ingels and J. Baldridge. Washington, DC: 1995.

Procedures Guide for Transcript Studies, NCES Working Paper 99–05, by M.N. Alt and D. Bradby. Washington, DC: 1999.

Survey Design

High School and Beyond First Follow-Up (1982) Sample Design Report, by R.E. Tourangeau, H. McWilliams, C. Jones, M.R. Frankel, and F. O'Brien. Washington, DC: 1983.

High School and Beyond Sample Design Report, by M. Frankel, L. Kohnke, D. Buonanno, and R. Tourangeau. Washington, DC: 1981.

High School and Beyond Second Follow-Up (1984) Sample Design Report, by C. Jones and B.D. Spencer. Washington, DC: 1985.

High School and Beyond Third Follow-Up Sample Design Report, CS 88–402, by B.D. Spencer, P. Sebring, B. Campbell, and D. Carroll. Washington, DC: 1987.

Psychometric Analysis of the NLS-72 and the High School and Beyond Test Batteries, by D.A. Rock, T.L. Hilton, J.M. Pollack, R.B Ekstrom, and M.E. Goertz. Washington, DC: 1985.

Data Quality and Comparability

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, and M. Saba. Washington, DC: 1997.

Quality of Responses of High School Students to Questionnaire Items, by W.B. Fetters, P. Stowe, and J.A. Owings. Washington, DC: 1984.

Chapter 9: SASS School Library Survey (SLS)

1. OVERVIEW

Federal surveys of school library media centers in elementary and secondary schools in the United States were conducted in 1958, 1962, 1974, 1978, and 1985. NCES now asks questions on libraries in public, private, and Bureau of Indian Affairs (BIA) schools as part of the Schools and Staffing Survey (SASS, see chapter 4). The School Library Media Center Survey was introduced as a component of SASS in 1993–94. It is sponsored by NCES and administered by the U.S. Bureau of the Census.

SAMPLE SURVEY OF ELEMENTARY AND SECONDARY SCHOOL LIBRARIES

SLS collects data on:
- Collections
- Expenditures
- Technology
- Services

Purpose

To provide a national picture of school library collections, expenditures, technology, and services. SLS furnishes national estimates for public and private school libraries (by school grade level and urbanicity) and for libraries operated by the Bureau of Indian Affairs (BIA) schools; state estimates for public schools; and national estimates for private school libraries, by detailed association. In 1993–94, SLS also furnished national and state estimates for public school librarians and estimates for private school librarians at the national level and by private affiliation or type of school.

Components

Before the School Library Media Center Survey was introduced in the 1993–94 SASS, questions on school libraries were asked in three components of the 1990–91 SASS. The *School Questionnaire* included items on the number of students served and the number of professional staff and aides. The *Teacher Demand and Shortage Questionnaire* included, at the district level, items on the number of full-time equivalent librarians/media specialists, vacant positions, positions abolished, and approved positions; and the *School Administrator Questionnaire* included items on the amount of librarian input in establishing curriculum.

The 1993–94 SLS component consisted of two questionnaires, one on the school's library media center and the other on the library media specialist. The 1999–2000 SASS included only the Library Media Center questionnaire. The surveys are sent to public schools, private schools, and BIA schools in the 50 states and the District of Columbia.

School Library Media Center Survey. The "*Library Survey*" is designed to provide a national picture of school library media center facilities, collections, equipment, technology, staffing, income, expenditure, and services. The respondents to the Library Survey are school librarians or other school staff members familiar with the library.

School Library Media Specialist/Librarian Survey. The "*Librarian Survey*" is designed to profile the school library media specialist workforce, including demographic characteristics, academic background, workload, career histories and plans, compensation, and perceptions of the school library media specialist profession and workplace. The eligible respondent for the Librarian Survey is the staff member whose main assignment at the school is to oversee the library.

Periodicity

The two surveys in SLS were first introduced in the SASS conducted during the 1993–94 school year. The Library Survey was repeated in the 1999–2000 SASS; the Librarian Survey was dropped from the 1999–2000 SASS.

2. USES OF DATA

School libraries and library media centers are an important component of the educational process. SLS data provide a national picture of school library collections, expenditures, technology, and services. The information can be used by federal, state, and local policymakers and practitioners to assess the status of school library media centers in the United States. It also contributes to the assessment of the federal role in supporting school libraries. The Librarian Survey provides, for the first time, a national profile of the school library media specialist/ librarian workforce.

SLS data can also be used to address current issues related to school libraries. Recent interest has focused on the contribution libraries could make to the current education reform movement. Education reform has prompted increased attention to the role school libraries/media centers might play in applying new technology and developing new teaching methods. Some analysts argue that libraries have a crucial role in developing computer literacy and educating students in the use of modern information technologies. A number of observers also have argued that expanding the function of libraries is a key prerequisite to meeting the National Education Goals.

3. KEY CONCEPTS

Some of the key concepts and terms in SLS are defined below. For additional terms, refer to the *1993–94 Schools and Staffing Survey: Data File User's Manual, Volume I: Survey Documentation* (NCES 96–142).

Librarian. A school staff member whose main responsibility is taking care of the library.

Library Media Center. An organized collection of printed, audiovisual, or computer resources that (a) is administered as a unit, (b) is located in a designated place or places, and (c) makes resources and services available to students, teachers, and administrators.

Library Media Specialist. A teacher who is state-certified in the field of library media.

4. SURVEY DESIGN

Target Population

The universe of library media centers/libraries and library media center specialists/librarians in elementary and secondary schools with any of grades 1–12 in the 50 states and the District of Columbia.

Sample Design

For the 1999–2000 SASS, the library media center sample was the entire SASS school sample, excluding charter schools. For more information on the 1999–2000 SLS sampling frame, refer to chapter 4, Schools and Staffing Survey (SASS). Each sampled library media center receives a library media center questionnaire.

In 1993–94, the library media center sample was a subsample of the SASS school sample. Drawn from the 13,000 schools in the SASS, the library sample consisted of 5,000 public schools, 2,500 private schools, and the 176 BIA schools in the United States. The librarian questionnaire was given to the head librarian of each sample library. (Thus, within a school, no librarian sampling took place.) The same strata were used for library sampling as were used for public school sampling (state and grade level). All BIA schools were selected for the library survey, so no stratification or sorting was needed. Within strata, public schools were sorted on the following variables: (1) LEA metro status (1=Central city of a metropolitan statistical area (MSA); 2=MSA, not central city; 3=Outside MSA); (2) LEA CCD ID; (3) school enrollment; and (4) school CCD ID.

SASS sample schools were then systematically subsampled using a probability proportionate to size algorithm, where the measure of size was the square root of the number of teachers in the school as reported in the Common Core of Data (CCD, the public school sampling frame for

SASS) times the school's inverse of the probability of selection from the public school sample file. Any school with a measure of size larger than the sampling interval was excluded from the library sampling operation and included in the sample with certainty.

The SASS private school library frame was identical to the frame used for the SASS private school survey, except that schools with special program emphasis, special education, vocational, or alternative curriculum were excluded. Private schools were stratified by recoded affiliation (Catholic, other religious, nonsectarian); grade level (elementary, secondary, combined); and urbanicity (urban, suburban, rural). Within each stratum, sorting occurs on the following variables: (1) Frame (list frame and area frame) and (2) school enrollment.

Within each stratum, schools were systematically selected using a probability proportionate to size algorithm. The measure of size used the school's measure of size times the school's inverse of the probability of selection. Any library with a measure of size larger than the sampling interval was excluded from the probability sampling process and included in the sample with certainty. In all, 2,500 private schools were selected for the library sample.

Data Collection and Processing

The U.S. Bureau of the Census is the collection agent for SLS. Data collection and processing procedures are discussed below.

Reference dates. Most data items refer to the most recent full week in the current school year. Questions on collections and expenditures refer to the previous school year.

Data collection. The Library Survey and, in 1993–94, the Librarian Survey are mailed with other components during October of the SASS survey year. The Library Surveys are addressed to "Principal" (and the 1993–94 Librarian Surveys were addressed to "Library Media Specialist/Librarian"). The follow-up procedures are described in chapter 4.

Editing. Once data collection is complete, data records are processed through a clerical edit, preliminary ISR classification, computer pre-edit, range check, consistency edit, and blanking edit. (See chapter 4 for details.) After the completion of these edits, records are processed through an edit to make a final determination of whether the case is eligible for the survey and, if so, whether sufficient data has been collected for the case to be classified as an interview. A final interview status code (ISR) value is assigned to each case as a result of the edit.

Estimation Methods

Weighting. Estimates from the SASS sample data are produced by using weights. The weighting process for each component of SASS includes adjustment for nonresponse using respondents' data, and—in 1993–94—adjustment of the sample totals to the frame totals to reduce sampling variability. Thus, weights for library sample schools that reported having a library were ratio adjusted to total SASS sample schools that reported having a library. Library sample schools that reported not having a library were similarly adjusted to study the characteristics of such schools. In the same fashion, library sample schools that reported having a librarian were ratio adjusted to total SASS sample schools that reported having a librarian, and library sample schools that reported not having a librarian were adjusted to study the characteristics of such schools. Due to reporting inconsistencies between the Library and Librarian Surveys and the School Survey, Library Survey data were not adjusted directly to schools reporting to have libraries, and Librarian Survey data were not adjusted directly to schools reporting to have librarians. The exact formula representing the construction of the weight for each component of the 1993–94 SASS is provided in the *1993–94 Schools and Staffing Survey: Sample Design and Estimation* (NCES 96–089).

Imputation. All item missing values are imputed for records classified as interviews. SLS uses a two-stage imputation procedure. In the first stage, items with missing values are completed whenever possible by using information about the school library/librarian from the following sources:

(1) Other questionnaire items on the same questionnaire;

(2) The matching Library Media Center (or Library Media Specialist/Librarian) Questionnaire; and

(3) The matching SASS School Questionnaire.

In general, the second stage of imputation fills remaining unanswered items by using data from the record for a library of a similar school; that is, a school that was the same level, of similar size, located in the same type of community, etc. Variables that describe certain characteristics of the schools (e.g., enrollment size and instructional level) are copied from the matching school record. In addition, a variable that categorizes the size of the library is created by using the number of books held at the end of the previous school year. These school variables and the library variable are used to sort the library records and to match incomplete records to those with complete entries (donors).

For some items, data are directly copied to the record with the missing value. For others, however, entries on the donor record are used as factors along with other information on the incomplete record to fill the items with missing values. For example, if the number of subscriptions acquired are reported for Library#1 but the number held is not, the donor's ratio of subscriptions held to subscriptions acquired is used with the number of subscriptions acquired by Library#1 to impute the number held by Library#1.

Remaining items with missing values are clerically imputed.

Recent Changes

The Librarian/Media Specialist component was not fielded in 1999–2000.

Future Plans

SASS administrations are now scheduled on a 4-year cycle. The next administration will be in 2003–2004.

5. DATA QUALITY AND COMPARABILITY

Although data are imputed for nonrespondents, caution should be exercised when analyzing data by state, sector, or affiliation. Since nonresponse varies by state, the reliability of state estimates and comparisons are affected. Users should be especially cautious about using data at a level of detail where the nonresponse rate is 30 percent or greater. See below for more information on types of error affecting data quality and comparability.

Sampling Error

The estimators of sampling variances for SASS statistics take the SASS complex sample design into account. See chapter 4.

Nonsampling Error

Nonresponse error.

Unit nonresponse. Data from the 1999–2000 Library Survey are not yet available. Weighted response rates for the 1993–94 Library Survey were 90.1, 70.7, and 89.4 percent for public, private, and BIA schools, respectively. Weighted response rates for the 1993–94 Librarian Survey were 92.3, 76.5, and 88.3 percent for the public, private, and BIA school librarians, respectively.

Item nonresponse. In 1993–94, several items had unweighted response rates below 75 percent in at least one of the public, private, or BIA versions of the survey. In the Library Survey, low-response items included questions on other audio-visual materials acquired by the library during school year; current serial subscriptions held at end of school year; other audio-visual materials held at end of school year, other audio-visual materials locally budgeted expenditures; video materials (tape & disc) locally budgeted expenditures; and number of students per week using the library media center. In the Librarian Survey, low-response items included field of study and year of doctorate or first professional degree; eight items on frequency of working with classroom teachers in the subject areas of reading, math, foreign language, etc.; two items on field of study and year of education specialist or professional diploma; and an item on whether the librarian was working in the school on a contributed service basis (private schools only).

Measurement error. A reinterview was conducted for the 1993–94 Library Survey. The library reinterview questionnaire collected information on 1993–94 library media center staffing, 1992–93 collection and expenditures, technology, library media center facilities, and scheduling and transactions. Full results from the reinterview study can be found in *Reinterview Report: Response Variance in the 1993 Library Survey.*

The reinterview was designed so that the data collection method was the same as that used in the original interview. For example, if the original interview was completed by mail, reinterview data was also collected by mail. If the original interview was completed by CATI (Computer Assisted Telephone Interview), the reinterview was done by CATI. For both methods of reinterview, the Census Bureau attempted to reinterview the same respondent who completed the original interview.

6. CONTACT INFORMATION

For content information on SLS, contact:

Jeffrey Williams
Phone: (202) 502–7476
E-mail: jeffrey.williams@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

1993–94 Schools and Staffing Survey: Data File User's Manual, Volume I: Survey Documentation, NCES 96–142, by K. Gruber, C.L. Rohr, and S.E. Fondelier. Washington, DC: 1996.

Uses of Data

Evaluation of Definitions and Analysis of Comparative Data for the School Library Statistics Program, NCES 98–267, by G. Dickson, Washington, DC: 1998.

Survey Design

1993–94 Schools and Staffing Survey: Sample Design and Estimation, NCES 96–089, by R. Abramson, C. Cole, S. Fondelier, B. Jackson, R. Parmer, and S. Kaufman. Washington, DC: 1996.

Data Quality and Comparability

Reinterview Report: Response Variance in the 1993 Library Survey, by P.J. Feindt. United States Department of Commerce, Bureau of the Census. Washington, DC: 1996.

Chapter 10: Public Libraries Survey (PLS)

1. OVERVIEW

The Public Libraries Survey (PLS) is the only source of current, national descriptive data on the status of public libraries in the United States. PLS is conducted annually by NCES through the Federal-State Cooperative System (FSCS) for Public Library Data. FSCS is a working network, allowing for close communication with the states through State Data Coordinators appointed by the Chief Officers of State Library Agencies (COSLA). At the federal level, NCES provides the financial support for FSCS activities. PLS data have been collected electronically by the U.S. Census Bureau, the collection agent for the PLS, since the first survey in 1989.

Purpose

To annually collect and disseminate descriptive data on all public libraries in the United States, the District of Columbia, and outlying areas, for use in planning, evaluation, research, and policymaking.

Components

There is one component to PLS. State Data Coordinators collect data from public libraries in their state, the District of Columbia, or outlying area and submit the completed survey to the U.S. Census Bureau. Outlying areas comprise the Commonwealth of the Northern Mariana Islands, Guam, Puerto Rico, the Republic of Palau, the U.S. Virgin Islands, and American Samoa.

Public Libraries Survey. Basic data items include the library's population of legal service area, full-time equivalent paid staff, service outlets, library materials, operating income and expenditures, capital outlay, circulation, reference transactions, library visits, public service hours, interlibrary loans, circulation of children's materials, children's program attendance, and as of 1995, interlibrary relationship, type of governance, administrative structure, several electronic measures, and whether or not the library meets all criteria of the FSCS definition of a public library. Identification items for public libraries include the library's name, address, telephone number, and county.

The same identification information is collected for public library service outlets and state library agencies. PLS also collects the following descriptive data on public library outlets and state library outlets: type of outlet, metropolitan status, number of books-by-mail-only outlets, web address, and number of bookmobiles. Four additional items are collected on characteristics of the state data submission: starting and ending dates for the fiscal year reporting period, official state total population estimate, and total unduplicated population of legal service areas.

Periodicity

Annual. Data are submitted for the previous fiscal year. The first PLS was for fiscal year 1989.

ANNUAL SURVEY OF THE UNIVERSE OF PUBLIC LIBRARIES

PLS collects data on:

- Population of legal service area
- Library staffing
- Operating income and expenditures
- Library materials
- Circulation, loan, and reference transactions
- Children's program attendance
- Electronic services
- Public service hours and visits

2. USES OF DATA

PLS provides the only current, national descriptive data on the status of nearly 9,000 public libraries. These data are used by federal, state, and local officials, professional associations, and local practitioners for planning, evaluation, and policymaking. Such valid, reliable, and timely statistics are essential for determining the investment of public resources in library development and operations. PLS data are also available to researchers and educators interested in issues related to public libraries. Because PLS is a universe that includes key characteristics such as legal basis (municipality, county, etc.) and location (urban, suburban, rural), it makes an excellent frame for drawing samples to address topics such as literacy, access for the disabled, library construction, electronic access, and services to children and young adults.

The FSCS Steering Committee and NCES foster the use and analysis of PLS data through annual training opportunities for State Data Coordinators. A Data Use Subcommittee addresses the dissemination, use, and analysis of PLS data.

3. KEY CONCEPTS

PLS collects identifying information on administrative entities and public library service outlets. An *administrative entity* is the public library, state library agency, system, federation, or cooperative service that is legally established under local or state law to provide public library service to a particular client group (e.g., the population of a local jurisdiction, the population of a state, or the public libraries located in a particular region). The entity may be administrative only and have no public library service outlets, have a single outlet, or have more than one outlet. The various administrative structures of public libraries are defined below. For other key terms, refer to the database documentation.

Public Library. Defined by FSCS as an entity established under state enabling laws or regulations to serve residents of a community, district, or region, and meeting these criteria: (1) has an organized collection of printed or other library materials, or a combination thereof; (2) employs a paid staff to provide and interpret such materials as required to meet the informational, cultural, recreational, and/or educational needs of a clientele; (3) has an established schedule in which services of the staff are available to the public; (4) has the facilities necessary to support such a collection, staff, and schedule; and (5) is supported in whole or in part with public funds. However, for purposes of the PLS data collection, state law prevails in the determination of a public library, and not all states define public libraries according to the PLS definition.

State Library Agency. The agency within each of the states and outlying areas which administers federal funds under the Library Services and Technology Act (LSTA) and is authorized to develop library services in the state or outlying area. It may also provide direct services to the public. Some state library agencies have service outlets.

System, Federation, or Cooperative Service. An autonomous library joined by formal or informal agreement(s) with other autonomous libraries to perform various services cooperatively, such as resource sharing and communications. In PLS, a public library may have the word "system" in its legal name but only identifies itself as a headquarters or member of a system, federation, or cooperative service if it has an agreement with another autonomous library. These agreements can be with other public libraries or with other types of libraries, such as school or academic libraries. Although data for library systems, federations, or cooperative services are not collected by PLS, the survey item "Interlibrary Relationship Code" indicates the system status of each public library.

Public Library Service Outlet. An outlet providing direct public library service and classified as one of the following types: central library outlet, branch library outlet, bookmobile outlet, or books-by-mail-only outlet. A public library may have one or more outlets, or it may have none.

Population of the Legal Service Area. The number of people in the geographic area for which a public library has been established to offer services and from which (or on behalf of which) the library derives income, plus any areas served under contract for which the library is the primary service provider. (Note that the determination of this population figure is the responsibility of the state library agency. The population figure should be based on the most recent official state population figures for jurisdictions in the state, available from the State Data Center. The State Data Coordinator obtains these figures annually from the State Data Center or other official state sources. For administrative entities that do not serve the public directly and have no outlets—e.g., a system, federation, or cooperative service—this number is zero. Population of the legal service area is a key survey item.)

4. SURVEY DESIGN

Target Population

All public libraries identified by the state library agencies in the 50 states and the District of Columbia, as well as libraries in outlying areas (Commonwealth of the Northern Mariana Islands, Guam, Puerto Rico, Republic of Palau, U.S. Virgin Islands, and American Samoa). Although data are not systematically collected from public libraries on Native American reservations, a category for Native American Tribal Government has been included in the survey item on type of local government structure since 1993. Data are not collected from military libraries that provide public library services or from libraries that serve residents of institutions.

Sample Design

PLS surveys the universe of public libraries.

Data Collection and Processing

PLS was the first national NCES survey in which respondents supplied the data electronically and in which data were edited and tabulated completely in machine-readable form. The states can submit their data by mail on diskette or over the Internet. The survey is generally released to the states over the Internet in the fall of the survey year, with returns due in the spring or summer (due date varies based on state fiscal cycle). Nonresponse follow up is conducted shortly thereafter.

Reference dates. The PLS reporting period is the previous fiscal year. If the fiscal year varies by locality, the state is requested to provide the earliest starting date and latest ending date reported by its public libraries. The last day of the fiscal year is the reference date for data on paid staff.

Data collection. As of fiscal year (FY) 98, states report their data using a personal computer Windows-based data collection software program which is downloaded from the Internet or available upon request on compact disc.

State level. The survey software has an edit check program that generates on-screen warnings during the data entry/import process, enabling respondents to review their data and correct many errors immediately. Following data entry/import, respondents can generate an on-screen or printed edit report for further review and correction of their data before submitting the final file to NCES. Four types of edit checks were performed: relational edit checks; out-of-range edit checks; arithmetic edit checks; and blank, zero, or invalid data edit checks.

Respondents also use the survey software to generate state summary tables and single-library tables (showing data for individual public libraries in their state). States are encouraged to review the tables for data quality before submitting their data to NCES. States submit their data with a signed form from the Chief Officer of the State Library Agency certifying its accuracy.

National level. NCES and the U.S. Bureau of the Census (the data collection agent for the survey) edit the state data submissions, working closely with the State Data Coordinators and the FSCS Steering Committee.

Estimation Methods

Imputation for nonresponding libraries was implemented with the 1995 PLS. FY 92 to FY 94 files were back-imputed for a 5-year trend report, which was released in 2001.

Imputation. Imputation was first implemented in 1995, using an imputation methodology developed by the Census Bureau. Annual public service hours were not imputed in 1995 but were imputed in later PLS cycles.

For many variables—such as numbers of audio books, bookmobiles, book/serial volumes, central, branches, librarians, reference transactions, etc.—data were imputed for nonresponding libraries categorized into imputation cells using a method which can be described as "updated cold deck"; that is, prior year's data were adjusted to accommodate the changes taking place over time. In some cases, prior year's ratios were applied to this year's data to impute some variables. For benefit and expenditure variables, logical procedures were used to impute the values; in some cases, a combination of the above methods were used. For libraries that did not respond for 2 years prior to the current survey, the mean value of an imputation cell was adjusted for a size variable of the missing units in the cell. For all nonresponding libraries, capital outlay was imputed by using expenditure variables and adjusting them when necessary.

Recent Changes

In 1995, imputation was implemented to compensate for nonresponse, and seven data items were added to the survey instrument. One new item asked whether or not the public library meets all criteria of the FSCS public library definition. The other items pertain to electronic technology, covering access to the Internet and electronic services, Internet usage, availability of library materials in electronic format, operating expenditures for electronic access, and expenditures for library materials in electronic format. New data elements added in 1998 were the number of Internet terminals used by staff only, and the number of Internet terminals used by the general public; deleted in 1998 on the Outlet file was the item on the population of legal service area by type of outlet, as the data were unreliable.

Future Plans

Web-based data collection is being considered for future surveys. NCES is developing a public library geographic mapping tool to be available on the Internet as part of the NCES Decennial Census School District 2000 project. This tool is an interactive online mapping system which integrates 2000 Decennial Census Data with school district boundaries and school district data. The library part of this tool will be developed in phases over the next several years.

5. DATA QUALITY AND COMPARABILITY

Data for nonresponding libraries were imputed beginning with the FY 95 survey. Before FY 95, the data were based on responding libraries only, and the percentage of public libraries responding to a given item varied widely among states. Therefore, *caution should be used in comparing FY 95 or later data to earlier data.* (Note: Imputed files have been produced for FY 92 to FY 94.)

State data comparisons should be made with caution because of differences in reporting periods and adherence to survey definitions. FSCS has formed a Definitions Subcommittee to work with the states on consistency of definitions and a Training Subcommittee to respond to the needs of the State Data Coordinators. *Special care should be used in comparing data for the District of Columbia, a city, with state data, and caution should also be used in making comparisons with the state of Hawaii, as Hawaii reports only one public library for the state.*

Public library questions are being included in other NCES surveys, including the National Household Education Surveys (NHES) and the Early Childhood Longitudinal Survey. Studies have been conducted to evaluate coverage, definitions, finance data, and staffing data. NCES has also sponsored a project to develop the first indices of inflation for public libraries, a cost index, and a price index, and another project that uses geographic mapping software to link census demographic data with PLS data. Work is under way to geocode public library service outlets nationwide and to map and digitize the boundaries of the nearly 9,000 public library legal service area jurisdictions so that they can be matched to Census Tiger files and to PLS data files.

Sampling Error

PLS is a universe survey and, therefore, not subject to sampling error.

Nonsampling Error

Differences in coverage from state to state, as well as differences in state laws and reporting practices, are the primary sources of nonsampling error in PLS.

Coverage error. The usage of different definitions of a public library may result in coverage error in some states. (See *Public Library Structure and Organization in the United States,* NCES 96-229.) Also, some outlying areas either do not submit the requested data or submit incomplete data; for this reason, not all outlying areas have been included in the data file or reports in past years. The Northern Marianas was included in both for the first time in FY 97, Guam in FY 98, and the Republic of Palau and the Virgin Islands in FY 2000.

In 1994, the Census Bureau conducted an evaluation of public library coverage in the 1991 PLS. (See *Report on Coverage Evaluation in the Public Library Statistics Program,* NCES 94-430.) This study showed PLS coverage to be very comprehensive, with only minor instances of undercounts or overcounts. The number of public libraries in the 1991 PLS relative to the number in state library directories was used as the measure of aggregate coverage. The coverage rate was 99.5 percent for the United States as a whole, and 87.5–106.3 percent for individual states. Thirty states had 100 percent coverage. The primary cause of undercoverage was nonresponse from some communities to their state's annual reporting requirement. Some of these states then excluded these communities' libraries from PLS.

Nonresponse error.

Unit nonresponse. The response rate to PLS is generally in the range of 97 to 99 percent. The response rate in 2000 was 98.3. The unit of response is the public library administrative entity that reports at least three of five key survey items (total paid employees, total income, total operating expenditures, book/serial volumes, and total

BEST COPY AVAILABLE

circulation), and that also reports population of the legal service area (provided by the State Data Coordinator). All 50 states and the District of Columbia have submitted data annually since the first survey in 1989. Six outlying areas added to PLS in 1993, but nonresponse or edit follow-up problems meant they were not included immediately in the data file or reports. The Northern Marianas was included for the first time in FY 97, Guam in FY 98, and the Republic of Palau and the Virgin Islands in FY 2000.

Item nonresponse. Response is generally 70 percent or higher for all items at the national level, but sometimes lower at the state level. In the FY 2000 PLS, response rates fell below 70 percent in several states for one or more of the following items: library visits, reference transactions, other income, total income, employee benefits, capital outlay, materials in electronic format, expenditures for materials in electronic format, Internet terminals used by staff only, audio materials, and users of electronic resources.

Measurement error. Several types of measurement error have been identified, largely related to inconsistencies in definitions used by the states and differences in their reporting practices.

Reporting period differences. The PLS reporting period is the previous fiscal year. There were eight different reporting periods in FY 2000, although most states reported data for the 12-month period of July to June or January to December. Fiscal year reporting may also vary by locality within a state; in such cases, the state is requested to provide the earliest starting date and latest ending date reported by its public libraries. While a state's reporting period may span more than a 12-month period, each library reports data for only a 12-month period.

Definitional differences. Definitions used by states in collecting data from their public libraries are not always consistent with PLS definitions. Three reports that address definitional problems are: ***Report on Evaluation of Definitions Used in the Public Library Statistics Program*** (NCES 95-430); ***Public Library Structure and Organization in the United States*** (NCES 96-229); and ***Report on Coverage Evaluation in the Public Library Statistics Program*** (NCES 94-430). The Definitions Subcommittee of the FSCS Steering Committee is working with the states to resolve these inconsistencies.

Estimates versus counts. Public libraries provide annual counts of library visits and reference transactions when counts are available. Otherwise, annual estimates are provided, based on a count taken during a typical week in October, multiplied by 52.

Population counts. There are significant methodological differences in the ways states calculate the three data items on population: (1) population of the legal service area of each public library administrative entity, (2) the total unduplicated population of legal service areas in the state, and (3) the official state total population estimate. There may also be differences in the time period for which the population data are provided. In addition, the calculated total for population of legal service areas of public libraries in a state sometimes exceeds the state's actual population or the state's total unduplicated population of legal service areas. This occurs when a state has overlapping service areas; that is, when adjacent libraries serve and thus count the same population.

6. CONTACT INFORMATION

For content information on public library statistics, contact:

Adrienne Chute
Phone: (202) 502–7328
E-mail: adrienne.chute@ed.gov

Elaine Kroe
Phone: (202) 502–7379
E-mail: patricia.kroe@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

Methodology discussed in Technical Notes.

General

Public Libraries in the United States: Fiscal Year 1999, NCES 2002–308, by A. Chute, E. Kroe, P. Garner, M. Polcari, and C.J. Ramsey. Washington, DC: 2002.

Public Libraries in the United States: Fiscal Year 2000, NCES 2002–344, by A. Chute, E. Kroe, P. Garner, M. Polcari, and C.J. Ramsey. Washington, DC: 2002.

Public Library Structure and Organization in the United States, NCES 96–229, by C. Kindel. Washington, DC: 1996.

Uses of Data

Finance Data in the Public Library Statistics Program: Definitions, Internal Consistency, and Comparisons to Secondary Sources, NCES 95–209, by C. Kindel. Washington, DC: 1995.

Measuring Inflation in Public Libraries: A Comparison of Two Approaches, the Input Cost Index and the Cost of Services Index, NCES 1999–326, by J.C. Chalmers and R. Vergun. Washington, DC: 1999.

Staffing Data in the Public Library Statistics Program: Definitions, Internal Consistency, and Comparisons to Secondary Sources, NCES 95–186, by C. Kindel and U.S. Bureau of the Census, Governments Division. Washington, DC: 1995.

Data Quality and Comparability

Data Comparability and Public Policy: New Interest in Public Library Data. Papers presented at Meetings of the American Statistical Association. NCES Working Paper 94–07. Washington, DC: 1994.

Report on Coverage Evaluation in the Public Library Statistics Program, NCES 94–430, by C. Kindel. Washington, DC: 1994.

Report on Evaluation of Definitions Used in the Public Library Statistics Program, NCES 95–430, by C. Kindel. Washington, DC: 1995.

Chapter 11: Academic Libraries Survey (ALS)

1. OVERVIEW

The Academic Libraries Survey (ALS) is designed to provide concise information on library resources, services, and expenditures for all academic libraries in the 50 states, the District of Columbia, and outlying areas. In 1998, ALS collected data on the approximately 3,650 libraries in the universe of higher education institutions. In the aggregate, these data provided an overview of the status of academic libraries nationally and statewide. The 1996 ALS also surveyed libraries in nonaccredited institutions that had a program of 4 years or more. Because so few of these libraries respond to ALS, their data were not published. Beginning with the 1998 ALS, the major distinction is whether the library is part of a postsecondary institution that was or was not eligible for Title IV funds.

Although ALS was a component of the Integrated Postsecondary Education Data System (IPEDS) from 1988 through 1998, ALS is now an independent survey.

BIENNIAL SURVEY OF THE UNIVERSE OF LIBRARIES IN HIGHER EDUCATION INSTITUTIONS

ALS collects data on:
- Library staffing
- Operating expenditures
- Total volumes
- Circulation, loan, and reference transactions
- Electronic services
- Gate count

Purpose

To periodically collect and disseminate descriptive data on all postsecondary academic libraries in the United States, the District of Columbia, and outlying areas, for use in planning, evaluation, and policymaking.

Components

There is a single component to the Academic Libraries Survey. The survey is completed by a designated respondent at the library. While ALS was a part of IPEDS, an appointed State IPEDS Data Coordinator collected the information from academic librarians and submitted it to NCES.

Academic Libraries Survey. Through 1996, ALS distinguished between libraries in postsecondary institutions accredited by agencies recognized by the Secretary of the U.S. Department of Education and libraries in nonaccredited institutions that had programs of 4 or more years. Starting with the 1998 collection, the major distinction is whether the library is part of a postsecondary institution that was or was not eligible for Title IV funds. Data include number of libraries, branches, and service outlets; full-time equivalent library staff by sex and position; operating expenditures by purpose, including salaries and fringe benefits; total volumes held at the end of the fiscal year; circulation transactions, interlibrary loan transactions, and information services for the fiscal year; hours open, gate count, and reference transactions per typical week; and as of 1996, the availability of electronic services such as electronic catalogs of the library's holdings, electronic full text periodicals, Internet access and instruction on use, library reference services by e-mail, electronic document delivery to patron's account-address, computers and software for patron use, scanning equipment for patron use, and services to the institution's distance education students.

Periodicity

Biennial in even-numbered years since 1990; triennial from 1966 to 1988.

2. USES OF DATA

Effective planning for the development and use of library resources demands the availability of valid and reliable statistics on academic libraries. ALS provides a wealth of information on academic libraries. These data are used by federal program staff to address various policy issues, by state policymakers for planning and comparative analysis, and by institutional staff for planning and peer analysis. Specific uses are listed below:

- Congress uses ALS data to assess the impact of library grant programs, the need for revisions of existing legislation, and the allocation of funds.
- Federal agencies that administer library grants for collections development, resource sharing, and networking activities require ALS data for their evaluation of the condition of academic libraries.
- State education agencies (SEAs) use ALS data to make comparisons at the national, regional, and state levels.
- Accreditation review programs for academic institutions require current library statistical data in order to evaluate postsecondary education institutions, establish standards, and modify comparative norms for assessing the quality of programs.
- Library administrators, academic managers, and national postsecondary education policy planners need current data on new electronic technologies to assess the impact of rapid technological change on the collections, budgets, and staffs of academic libraries. College librarians and administrators need these data to develop plans for the most effective use of local, state, and federal funds. Staff data are input to supply/demand models for professional and paraprofessional librarians.
- Library associations—such as the American Library Association, the Association of Research Libraries, and the Association of College and Research Libraries—use ALS data to determine the general status of the profession. Other research organizations use the data for studies of libraries.
- Program staff in the Institute of Education Sciences of the U.S. Department of Education use ALS data for administering their library grants program, evaluating existing programs, and preparing documentation for congressional budget hearings and inquiries.

3. KEY CONCEPTS

Some of the key concepts and terms in ALS are defined below. For additional terms, refer to *Integrated Postsecondary Education Data System: Glossary* (NCES 97–543).

Academic Library. A library operated by a postsecondary education institution that has: (1) an organized collection of printed, microform, and audiovisual materials; (2) a staff trained to provide and interpret such materials as required to meet the informational, cultural, recreational, or educational needs of clientele; (3) an established schedule in which services of the staff are available to clientele; and (4) the physical facilities necessary to support such a collection, staff, and schedule. Units that are part of a learning resource center are included if they meet the above criteria.

Branch Library. An auxiliary library service outlet with quarters separate from the central library of an institution. A branch library has a basic collection of books and other materials, a regular staffing level, and an established schedule.

Volume. Any printed, mimeographed, or processed work, contained in one binding or portfolio, hardbound or paperbound, that has been catalogued, classified, or otherwise made ready for use.

Title. A publication that forms a separate bibliographic whole, whether issued in one or several volumes, reels, disks, slides, or parts. The term applies equally to printed materials (e.g., books and periodicals), sound recordings, film and video materials, microforms, and computer files.

Circulation Transaction. Includes all items lent from the general collection and from the reserve collection for use generally (although not always) outside the library. Includes both activities with initial charges (either manual or electronic) and renewals, each of which is reported as a circulation transaction.

Interlibrary Loan. A transaction in which library materials, or copies of the materials, are made available by one library to another upon request. Loans include providing materials and receiving materials. Libraries involved in these interlibrary loans cannot be under the same administration or on the same campus.

Reference Transaction. An information contact that involves the knowledge, use, recommendation, interpretation, or instruction in the use of one or more information sources by a member of the library staff. Information sources include printed and nonprinted materials, machine-readable databases (including assistance with computer searching), catalogues and other holdings records, and, through communication or referral, other libraries and institutions and persons both inside and outside the library. Includes information and referral services.

Online Public Access Catalogue (OPAC). A library's catalog of its collections in electronic form accessible by computer or other online workstation.

Gate Count. The total number of persons physically entering the library in a typical week.

4. SURVEY DESIGN

Target Population

The libraries of all institutions in the 50 states, the District of Columbia, and the outlying areas that have as their primary purpose the provision of postsecondary education. Branch campuses of U.S. institutions located in foreign countries are excluded. Through 1996, ALS distinguished between libraries in postsecondary institutions accredited by agencies recognized by the Secretary of the U.S. Department of Education and libraries in nonaccredited institutions that had programs of four or more years. In 1996, there were approximately 3,600 accredited institutions and 400 nonaccredited institutions in the IPEDS universe. About 3,400 of the accredited institutions had academic libraries. Starting with the 1998 collection, the major distinction is whether the library is part of a postsecondary institution that was or was not eligible for Title IV funds.

Sample Design

ALS surveys the universe of postsecondary institutions.

Data Collection and Processing

The 2000 ALS was a web collection. The U.S. Bureau of the Census is the collection agent. In recent administrations, State IPEDS Data Coordinators collected, edited, and submitted ALS data to the Census Bureau, using the software package IDEALS (i.e., Input and Data Editing for Academic Library Statistics). An academic librarian in the state assisted with the collection and submission of the data.

Reference dates. Most ALS data are reported for the most recent completed fiscal year, which generally ends before October 1 of the survey year. Information on staff and services per typical week are collected for a single point in time during the fall of the survey year, usually the institution's official fall reporting date or October 15.

Data collection. In the 2000 ALS web collection, library respondents submitted data directly to the Census Bureau through the web. Libraries began receiving registration materials in August and could submit responses from October through the following February. A web-based survey is the latest in a number of steps to improve ALS collection. In July 1990, NCES initiated an ALS improvement project with the assistance of the National Commission on Libraries and Information Science (NCLIS) and the American Library Association's Office of Research and Statistics (ALA-ORS). The project identified an academic librarian in each state to work with the IPEDS Coordinators in submitting their library data. During the 1990s, many of these library representatives took major responsibility for collecting data in their state. Others were available to assist in problem resolution when anomalies are discovered in completed questionnaires.

The ALS improvement project also led to the development of the microcomputer software package (IDEALS), which was used by states in reporting their academic library data. Along with the software, NCES provided IPEDS Data Coordinators with a list of instructions explaining precisely how responses were to be developed for each ALS item. Academic librarians within each state completed hard copy forms, as they had previously, and returned them to the state's library representative or IPEDS Coordinator. States were given the option of submitting the paper library forms but were encouraged to enter the data into IDEALS and submit the data on diskette to the Census Bureau. Nearly all states elected the diskette option.

ALS was mailed to postsecondary institutions during the summer of the survey year, with returns requested during the fall. Any survey returns from institutions that did not have an academic library were declared to be out of scope, as were institutions that did not have their own library but shared one with other institutions. In recent years, less than half of the nonaccredited institutions responded to the survey; NCES does not include data on this group in publications because the estimates are not statistically acceptable.

Editing. The web-based collection incorporates most of the internal consistency edit checks, range checks, and summation checks that the IDEALS software featured, but allows these checks to be run at the library level instead of at the state level. These edit checks provide some warning as the data are being keyed. When the

BEST COPY AVAILABLE

IDEALS software was used, library representatives at the state level could also run edit/error reports and make corrections before submitting the data to NCES. Examples of these edit checks include summation checks, relational edit checks, and range checks.

When probable errors are identified, Census Bureau personnel contact the institution to resolve the problem. After all the data are received, general edits are performed. These edits include checks for comparability between the response to the "own library inquiry" in ALS and the Institutional Characteristics Survey; between expenditures for staff reported in Part C of the ALS questionnaire and full-time equivalent staff reported in Part B; between expenditures on books, etc. in Part C and the numbers of books, etc. reported in Part D; between library holdings at the end of the year and the number of materials added during the year; between the number of presentations given and the number of persons served in presentations; and between the library data reported in the current survey and the same data reported in the prior survey. Once all edits have been performed and all corrections have been made, the data undergo imputation to compensate for nonresponse (see below).

Estimation Methods

Imputation is used in ALS to compensate for nonresponse. In 1994, procedures were changed to use data from the previous survey if available, and only use imputation group means (see below) if prior-year data were not available. Before 1994, only imputation group means were used.

Imputation. ALS imputation is based on the response in each part of the survey. Each part goes through either total or partial imputation procedures except Part A, Number of Branch and Independent Libraries; Part B, Line 4—Library staff information-contributed services staff; and Part C, Line 23—Library operating expenditures-employee fringe benefits. These items are imputed only if reported prior year data are available (contributed services staff and employee fringe benefits apply to only a few institutions). Part G, Electronic Services, does not go through imputation.

The imputation methods use either prior year data or current year imputation group means. The procedures are slightly different depending on whether an institution is totally nonresponding or partially nonresponding in the current year. If prior year data are available, the imputation procedure either carries forward the prior year data or carries forward the prior year data multiplied by a growth factor. If prior year data are not available, the imputation procedure uses the current year imputation group means as the imputed value.

Means and ratios are calculated for each of eight imputation groups. There are three imputation groups each for public, 4-year or above institutions and private, 4-year or above institutions: (a) those granting 50 or more doctoral degrees; (b) those granting less than 50 doctoral degrees and 50 or more postbaccalaureate degrees; and (c) all others. The remaining two imputation groups combine (1) public, 2-year institutions and public, less than 2-year institutions; and (2) private, nonprofit, 2-year institutions; private, for-profit, 2-year institutions; private, nonprofit, less than 2-year institutions; and private for-profit, less than 2-year institutions. Note that computation of the imputation base excludes institutions that merged, split, submitted combined forms, changed sectors from the prior year, or did not submit a full report for either the current year or the prior year.

Some examples follow:

If a total is blank or zero, but there are one or more positive subtotals, the total is changed to equal the sum of the subtotals. Alternatively, if, for a given record, there is a reported total but all subtotals are either zero or blank, then it is assumed that the subtotals should have positive values and values are imputed.

To calculate the imputed value for a subtotal, the average estimate is calculated across the set of respondents including ones for which the total is obtained by adding the subtotals, but excluding those for which the sum of the subtotals does not originally equal the total. The average subtotal value is divided by the average total value within each imputation group to obtain an average proportion. The average proportion is then multiplied by the reported total to obtain the imputed subtotal value.

For key items *total staff* and *total operating expenditures*, if the total and all subtotals are blank or zero, they are imputed by using the average by imputation group from the set of respondents described above. Zero is not a valid entry for these items.

The imputation procedures of using a ratio adjustment to prior year data for imputation represented a change from that followed in cycles prior to 1996, and may have resulted in some small differences in estimates. While checks indicate that the effect of the change in imputation procedures was not large, caution should be exercises in making comparisons with pre-1996 or earlier reports. See Status of Academic Libraries in the United States: Results from the 1996

Academic Library Survey with Historical Comparisons (NCES 2001–301).

Recent Changes

Several changes were made to the survey instrument in 1996, 1998, and 2000. These are summarized below. In the 1996 instrument, the data items in Part E of the questionnaire (Library Services) were expanded to request separate reporting for returnables and nonreturnables, as well as totals. In addition, a new section, Part G, was added to collect information about access to the following electronic services, both on and off campus:

- Electronic catalog that includes the library's holdings;
- Electronic indexes and reference tools;
- Electronic full text periodicals;
- Electronic full text course reserves;
- Electronic files other than the catalog (e.g., finding aids, indices, manuscripts) created by library staff;
- Internet access;
- Library reference service by e-mail;
- Capacity to place interlibrary loan/document delivery requests electronically;
- Electronic document delivery by the library to patron's account/address;
- Computers not dedicated to library functions for patron use inside the library;
- Computer software for patron use inside the library (e.g., word processing, spreadsheet, custom applications, etc.);
- Technology in the library to assist patrons with disabilities (e.g., TDD, specially equipped workstations); and
- Instruction by library staff on use of Internet resources.

The 1998 ALS survey instrument modifications included the following.

The definition of a library was moved to the cover page and reformatted as a checklist. The other cover page change was that the possibilities of reporting data *for* another library or having data reported *by* another library were clarified. The data items in Part B (Library Staff) were expanded to request a total full-time equivalency (FTE) count for librarians and other professionals as well as separate counts of these two categories of staff. Part C was renamed "Library Expenditures" and the word "operating" was used only in reference to expenditures for items other than staff and materials. The two major lines for reporting expenditures on information resources were subdivided as follows: books, serial backfiles, and other materials (paper and microform; electronic); and current serial subscriptions and search services (paper and microform; electronic). In addition, expenditures on search services were to be reported with those for current serial subscriptions, in recognition of the fact that it is often impossible to separate the two. Part D (Collections) was changed the most, being reduced from 18 lines to 7. It collected data on only three types of materials: books, serial backfiles, and other materials (paper; microform; electronic); current serial subscriptions (paper and microform; electronic); and audiovisual materials. The following lines were deleted: manuscripts and archives, cartographic materials, graphic materials, sound recordings, film and video materials, and computer files. Except for paper materials, there was no longer separate reporting of physical counts and title counts. In Part F (Library Services, Typical Week), "Public service hours" was changed to "hours open" since some libraries keep two separate counts and were unsure of what to report. "Typical week" was added to the heading above the space for reporting figures to reinforce that only typical week figures should be reported. In Part G (Electronic Services), the following items were added to the yes/no checklist about access to electronic services:

- Computers not dedicated to library functions for patron use inside the library;
- Computer software for patron use in the library (e.g., word processing, spreadsheet, custom applications, etc.);
- Scanning equipment for patron use in the library; and
- Services to your institution's distance education students.

The changes for the 1998 form for the 2000 ALS are as follows:

Cover sheet (Library Definition): The format of the question regarding providing financial support to another library was clarified.

Part C (Library Expenditures): The text for library expenditures was modified to clarify what is wanted.

Part D (Library Collections): The items "Electronic-Titles" and "Number of electronic subscriptions" were dropped and the item covering other forms of subscriptions was revised.

Part E (Library Services): A new item was added for "Documents delivered from commercial services" and the words "document delivery" were dropped from the items for "interlibrary loans provided" and "interlibrary loans received."

Part G (Electronic Services): Five items were added under the heading "Consortial Services."

Future Plans

At this time, NCES plans to continue conducting ALS biennially.

5. DATA QUALITY AND COMPARABILITY

NCES makes every effort to achieve high data quality. Through a web collection that includes built-in edit checks, it hopes to improve the quality of ALS data. *Users are cautioned about limitations in the analysis of ALS data by state or by level and control of institution. Since nonresponse varies by state, the reliability of state estimates and comparisons are affected. Special caution should be exercised when using data where the nonresponse rate is 30 percent or greater.* See below for more information on the types of error affecting data quality and comparability.

Sampling Error

Because ALS is a universe survey, there is no sampling error.

Nonsampling Error

Coverage error. A comprehensive evaluation of the coverage of ALS found that quality of institutional coverage was excellent (a coverage gap of only 1 to 3 percent) when compared to other institutional listings directly related to the academic libraries industry, although questions remain as to whether the data collected by ALS fully account for branch data associated with parent institution resources. (See *Coverage Evaluation of the Academic Library Survey*, NCES 1999-330.) A second problem plaguing ALS data is the presence or absence of professional school statistics in parent college or university data.

Nonresponse error.

Unit nonresponse. The overall unit response rate for the 1998 ALS was 97.0 percent, higher than in 1996 (94.2 percent) or 1994 (93.7 percent). Nineteen states had response rates of 100 percent, and 19 states fell below the overall rate of 97.0 percent; their rates ranged from 71.4 to 96.9 percent. The aggregate response rate for 4-year institutions was 97.7 percent (ranging from 97.0 percent for master's level to 98.8 percent for doctor's degree). Institutions of less than 4 years had a slightly lower response rate of 95.8 percent. Overall response rates were 98.2 percent for public institutions and 96.0 percent for private institutions.

Item nonresponse. In the 1998 ALS, 23 items had response rates of 90 percent or higher; 63 items had rates in the 80–89 percent range; 12 items had rates in the 70–79 percent range; and 4 items had rates lower than 70 percent. One of these items was in the area of library staff (69.5 percent), one in the area of library operating expenditures (66.0 percent), and two in the area of library collections (65.2 and 65.3 percent).

Measurement error. No information available.

6. CONTACT INFORMATION

For content information on ALS, contact:

Jeffrey Williams
Phone: (202) 502–7476
E-mail: jeffrey.williams@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Academic Libraries: 1998, NCES 2001–341, by M.W. Cahalan and N.M. Justh. Washington, DC: 2001.

Academic Libraries: 1996, NCES 2000–326, by M.W. Cahalan and N.M. Justh. Washington, DC: 2000.

Data Quality and Comparability

Coverage Evaluation of the Academic Library Survey, NCES 1999-330, by C.C. Marston. Washington, DC: 1999.

Status of Academic Libraries in the United States: Results from the 1996 Academic Library Survey with Historical Comparisons, NCES 2001–301, by M. Cahalan, W. Mansfield, and N. Justh. Washington, DC: 2001.

Chapter 12: State Library Agencies (StLA) Survey

1. OVERVIEW

The State Library Agency (StLA) Survey collects data annually on state library agencies in the 50 states and the District of Columbia. This survey is the product of a cooperative effort between the Chief Officers of State Library Agencies (COSLA), the National Commission on Libraries and Information Science (NCLIS), and NCES. The first StLA Survey collected data for fiscal year 1994.

Purpose

To provide descriptive information about all StLAs in the 50 states and the District of Columbia.

Components

There is one component to the StLA Survey. StLA staff collects the information.

StLA Survey. This survey collects data on governance, public service hours, number and types of service outlets, type and size of collections, library service transactions and development transactions, electronic services and information, resources assigned to allied operations (e.g., archive and records management), staffing, income, and expenditures. Data are also collected on StLA services to public, academic, school, and special libraries, and to library systems.

Periodicity

Annual. Data are submitted for the previous fiscal year. The first StLA Survey was for fiscal year (FY) 1994.

ANNUAL SURVEY OF THE UNIVERSE OF STATE LIBRARY AGENCIES

StLA collects data on:

- Governance
- Library staffing
- Income and expenditures
- Type and size of collections
- Service and development transactions
- Electronic services
- Public service hours
- Number and types of service outlets

2. USES OF DATA

The StLA Survey provides state and federal policymakers, researchers, and other interested users with a wealth of descriptive information about StLAs in the 50 states and the District of Columbia. It provides data on the variety of roles played by StLAs and the various combinations of fiscal, human, and informational resources invested in their work. Together with other NCES data collections on public, academic, school, and federal libraries, and on library cooperatives, the StLA Survey provides a comprehensive profile of libraries and information services in the United States.

3. KEY CONCEPTS

A few key concepts are defined below. For definitions of all terms, refer to the survey instrument in the database documentation.

BEST COPY AVAILABLE

State Library Agency (StLA). The official agency of a state that is (1) charged by the law of that state with the extension and development of public library services throughout the state, and (2) responsible for administering federal funds under the Library Services and Technology Act (LSTA), Public Law 104-208. Beyond these two essential roles, StLAs vary greatly. They can be located in different departments of state government and report to different authorities, are involved in various ways in the development and operation of electronic information networks, and provide different types of services to different types of libraries.

The administrative and developmental responsibilities of StLAs affect the operation of thousands of public, academic, school, and special libraries in the nation. StLAs also provide important reference and information services to their state government, and administer their state library and special operations such as the state archives, libraries for the blind and physically handicapped, and the State Center for the Book. An StLA may function as its state's public library at large, providing service to the general public and state government employees.

Academic Library. A library forming an integral part of a college, university, or other academic institution for postsecondary education, and organized and administered to meet the needs of students, faculty, and affiliated staff of the institution.

Public Library. A library that serves all residents of a given community, district, or region, and that typically receives its financial support, in whole or part, from public funds.

School Library Media Center. A library that is an integral part of the educational program of an elementary or secondary school, with materials and services that meet the curricular, information, and recreational needs of students, teachers, and administrators.

Special Library. A library in a business firm, professional association, government agency, or other organized group; a library that is maintained by a parent organization to serve a specialized clientele; or an independent library that may provide materials or services, or both, to the public, a segment of the public, or to other libraries. The scope of collections and services are limited to the subject interests of the host or parent institution. Includes libraries in state institutions (e.g., state-run prisons, hospitals, and residential training schools).

System. A group of autonomous libraries joined together by formal or informal agreements to perform various services cooperatively such as resource sharing, communications, etc. Includes multitype library systems and public library systems. Excludes multiple outlets under the same administration.

Allied Operations. Other information resources with which the StLA may be affiliated. Includes the state archives; state legislative reference/research service; state history museum/art gallery; and state records management service. Excludes the State Center for the Book and libraries for the blind and physically handicapped.

Collections. The volumes or physical units in all StLA outlets (main or central libraries, bookmobiles, and other outlets) that serve the general public and/or state government. Includes book and serial volumes (excluding microforms), audio materials, video materials, serial subscriptions, and government documents.

4. SURVEY DESIGN

Target Population

The state library agencies in the 50 states and the District of Columbia (51 total).

Sample Design

The StLA Survey covers the universe of state library agencies in the 50 states and the District of Columbia.

Data Collection and Processing

As of the FY 99 StLA Survey, NCES collects the data via an Internet web-based reporting system, as described below. (Prior to FY 99, the data were collected via customized survey software.) The web survey is usually released on the web in mid-October with a due date in mid-February. Nonresponse follow up is conducted immediately after receipt of the completed survey over the Internet. The U.S. Bureau of the Census serves as the data collection and processing agent for NCES.

Reference dates. The reporting period for the StLA Survey is the previous fiscal year. The reference date for reporting staff counts is October 1.

Data collection. Beginning in FY 99, the data are reported through an Internet web-based reporting system designed to reduce respondent burden and enable states to edit their data before submission to NCES. The system contains prior-year data for items where the data are not expected to change annually—about 40 percent of the survey items. The respondent is requested to review the pre-entered data and update any information that has changed. The respondent is instructed to answer all other items; to enter -1 to any numeric item if the data cannot be provided; and to report 0 if a count is taken with a result of zero. Items left blank indicate nonresponse (i.e., not reported or not applicable). Respondents are alerted to questionable data during the data entry process through interactive, on-screen error warnings that prompt them to verify or revise the data, as appropriate. The web-based system also provides error/warning reports of questionable data that can be reviewed on-screen or printed. These features allow the respondent to submit a data file that requires minimal or no follow up for data problems.

Editing. Data from the StLA Survey are edited by the states and NCES in different stages, based on established editing criteria.

State level. The web-based system performs four types of edit checks before the data are submitted to NCES: relational edit checks; out-of-range edit checks; arithmetic edit checks; and blank/zero/invalid edit checks.

National level. NCES, assisted by the Census Bureau, edits individual state submissions by e-mail and telephone follow-up with survey respondents. After submissions are received from all 50 states and the District of Columbia, the preliminary national file and draft tables for the *E.D. TABS: State Library Agencies* publication are reviewed for data quality by the StLA Steering Committee, NCES, and the Census Bureau. States with questionable data are contacted to request verification or correction of their data before the final file and tables are produced.

Estimation Methods

StLA began imputing for item nonresponse as of FY 99.

Imputation. Missing data are imputed using one of four methods, in the following order: the zero rule, the growth rule, regression modeling, or the sum rule. Under the zero rule, if the state does not report a value for the current year and reported zero for the prior year, then the value for the current year is set to zero. This rule is applied first, on the assumption that there was no change from the prior year. Under the growth rule, if the state does not report a value for the current year and the value for the prior year was greater than zero, the growth rate from the prior year to the current year is calculated for all states that reported data greater than zero in both years. The median of the growth rates is then calculated and applied to the state's previously reported data to obtain an estimate for the current year. (Note that the growth rule looked at values for the prior year only.) Regression modeling is used if the state does not report a value for the current year and there was no value for the prior year. The regression model uses only the current year's data file. It uses three to six auxiliary items reported by all states to determine the regression model that best fit the data. The auxiliary items are selected by calculating the correlations between the imputed item and all other numeric items on the data file, and, after a process of elimination, using the items that have the highest correlations to the imputed item. The sum rule applies when the details of a total and the total are missing, and the details are imputed by the zero rule, the growth rule, or regression modeling; the total is imputed by adding up the details.

Recent Changes

A number of changes were made to the 2002 survey, particularly to Part F-Electronic Services and Information. In Part D, the responses to all items in one question were revised to clarify how the StLA provided services. In Part E, one item was revised to indicate that only one StLA outlet may be identified as the main or central outlet, and another question was split into two to provide more information about hours open. In Part F, the Serial Subscription item was revised to clarify that only current serial subscriptions in print format should be reported. In Part N, one question was split into two to collect more specific information on Internet workstations owned by the StLA or available but not owned by the StLA, and another question was revised to include a new Bibliographic Records item. Two changes were made to a third question: an Other Expenditures item was added for consistency with items collected in Part K, and the OCLC Participation and Z39.50 Gateway items were deleted. Finally, two items were added to Part J to identify the types of libraries for which StLAs administer state funds, and six items were added to Part N to collect more current descriptive data on electronic services provided by StLAs.

Future Plans

No changes are currently planned for the FY 03 survey.

BEST COPY AVAILABLE

5. DATA QUALITY AND COMPARABILITY

Data from the StLA Survey were not imputed for item nonresponse prior to FY 99, so state and national totals for some items may be underestimated in earlier years. *State comparisons should be made with caution because item response rates, fiscal year reporting periods, and adherence to survey definitions vary by state. Special care should also be taken in comparing data for the District of Columbia (a city) with data for a state.*

Sampling Error

The StLA Survey is a universe survey and, therefore, not subject to sampling error.

Nonsampling Error

Coverage error. There is no coverage error in the StLA Survey. It includes the universe of state library agencies in the 50 states and the District of Columbia.

Nonresponse error.

Unit nonresponse. The StLA Survey has achieved a 100 percent response rate in all survey administrations.

Item nonresponse. Most items have a 100 percent response rate. In FY 01, only six items did not have a 100 percent response rate: five items had a response rate of 98.0, and one had a response rate of 88.2 percent.

Measurement error. Measurement (or reporting) errors can result from the use of different definitions for key terms and different reporting periods among the states. The fiscal year of most states is July 1 to June 30. Exceptions are New York (April 1 to March 31); Texas (September 1 to August 31); and Alabama, the District of Columbia, and Michigan (October 1 to September 30).

Some definitions of selected fiscal data related to the Library Services and Construction Act (LSCA), the predecessor to the LSTA, needed clarification, based on inconsistent reporting of the data. The Census Bureau conducted an evaluation study to examine these data, and the survey instructions for various LSCA items on income and expenditures were revised based on the report recommendations. Specifically, the instructions for the reporting of LSCA income and LSCA expenditures for statewide services and financial assistance to libraries and systems were clarified.

Although some data for two states should have been reported in the Public Libraries Survey (see chapter 10) instead of in the 1994 StLA Survey, NCES has negotiated successfully with these StLAs to eliminate such reporting from the 1995 and later StLA Surveys.

6. CONTACT INFORMATION

For content information on the StLA Survey, contact:

Elaine Kroe
Phone: (202) 502–7379
E-mail: patricia.kroe@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

Methodology discussed in technical notes to survey reports.

General

State Library Agencies, Fiscal Year 2001, NCES 2003–309, by B. Holton, E. Kroe, P. O'Shea, C. Sheckells, S. Dorinski, and M. Freeman. Washington, DC: 2002.

State Library Agencies, Fiscal Year 2000, NCES 2002–302, by E. Kroe, P. Garner, and C. Sheckells. Washington, DC: 2001.

State Library Agencies, Fiscal Year 1999, NCES 2000–374, by E. Kroe. Washington, DC: 2000.

State Library Agencies, Fiscal Year 1998, NCES 2000–318, by E. Kroe. Washington, DC: 2000.

State Library Agencies, Fiscal Year 1997, NCES 1999–304, by E. Kroe. Washington, DC: 1999.

Data Quality and Comparability

Evaluation of the NCES State Library Agencies Survey: An Examination of Duplication and Definitions in the Fiscal Section of the State Library Agencies Survey, NCES 1999–312, by L.R. Aneckstein. Washington, DC: 1999.

Chapter 13: Federal Libraries and Information Centers Survey

1. OVERVIEW

Since 1965, NCES has periodically conducted a comprehensive survey of federal libraries in the 50 states and the District of Columbia. The 1994 Federal Libraries and Information Centers Survey was the sixth survey, the first since 1978, and the first to include information centers. This survey is a cooperative effort of the National Center for Education Statistics (NCES) and the Federal Library and Information Center Committee (FLICC) of the Library of Congress. There are no current plans for the next administration of the survey.

PERIODIC SURVEY OF THE UNIVERSE OF FEDERAL LIBRARIES

Collects data on:
- Library staffing
- Library collections
- Service per typical week
- Automation and technology
- Preservation

Purpose
To provide descriptive information about all federal libraries and information centers in the 50 states and the District of Columbia, excluding elementary and secondary school libraries under federal agency operation.

Components
There is only one component to the Federal Libraries and Information Centers Survey. The survey is completed by a designated respondent at the library or information center.

Federal Libraries and Information Centers Survey. This survey collects the following information on federal libraries and information centers: staffing, collections, service per typical week, automation, technology, and preservation.

Periodicity
Irregular. The survey previous to the 1994 survey was conducted in 1978, and there are no current plans for the next administration.

2. USES OF DATA

The 1994 Federal Libraries and Information Centers Survey updates the federal library survey data collected in 1978, establishing a more current national profile of federal libraries and information centers. A primary use of this survey's data is the publication of the *Directory of Federal Libraries and Information Centers*, which provides for each entry the name, address, and type of library or information center, and the name and telephone number of a contact person. The type of library or information center represents the library/information center's primary subject-matter acquisitions, categorized as follows: presidential, national, academic, engineering and science, health and medicine, general, law, multitype, training center and/or instructional technical school, and special. Most of the information in the Directory is provided by survey respondents.

For nonrespondents, the name and address of the library or information center are obtained from the file used to conduct the survey. The latest Directory represents the universe of domestic federal libraries and information centers as of September 30, 1994. Changes available prior to publication were incorporated.

3. KEY CONCEPTS

The terms defined below are a subset of the terms in the Federal Libraries and Information Centers Survey. For definitions of all terms, refer to the survey instrument in the database documentation.

Library/Information Center. A *library* is an organization that includes among its functions the following: selection, acquisition, organization, preservation, retrieval, and provision of access to information resources. An *information center* is an organization that performs the function of linking requestors with appropriate information resources through established mechanisms, such as searching databases, providing referrals, answering specific questions, or by other means. A library or information center may be further defined as:

Autonomous. One that has a separate facility, collection, staff, defined clientele, and full operational control. The principal operating budget generally derives from the institution served.

Headquarters. Either a single-unit library serving administrative headquarters or a central user unit with administrative and directional control of other libraries.

Central/main. The single-unit library or the administrative center of a multi-unit library where the principal collections are kept and handled.

Branch or nonautonomous. A user-service unit which has all of the following:

- quarters that are separate from the central library;
- a permanent basic collection of material;
- a permanent staff provided by the central library or the institution or organization of which the library is a part; and
- a regular schedule for opening.

Such units are administered from the central library. Although they are not autonomous, some units may report independently for the purpose of this survey.

Network and Cooperative. Two or more independent libraries of any type(s) engaging in cooperative activities to perform library services for mutual benefit, according to some agreement on common purposes while retaining individual autonomy. The activities extend beyond reciprocal borrowing and beyond the scope of the national (American Library Association) interlibrary loan code.

Bibliographic Service Center. An organization that serves a network of libraries as a distributor of computer-based bibliographic services. A service center gains access to bibliographic data through a bibliographic utility.

Bibliographic Utility. An organization that maintains online databases provided by various libraries individually or cooperatively through networks. The utility provides a standard interface by which bibliographic data are accessible to libraries either directly or through bibliographic service centers.

Centralized Processing Center. A library or other agency that orders library materials, prepares these materials for use, and prepares cataloguing records for these materials on behalf of a group of libraries.

Cooperative Collection Resource Facility. A facility supported cooperatively by a group of libraries to acquire, maintain, and provide access to collection resources not generally available in any or all of the cooperating libraries. Materials may be acquired through cooperative purchase or through depository arrangements to maintain little-used materials furnished by participating libraries. Services typically include interlibrary lending, photocopying, and materials preservation. This type of facility is distinguished from a storage facility in which materials stored cooperatively remain the property of each library rather than becoming common property of the facility. The Center for Research Libraries is one example of a cooperative collection resource facility.

Gate Count. The number of persons counted either entering or leaving the library/information center in a typical week in the past year. If not regularly counted, results of samplings may be entered.

FEDLINK. A cooperative network program (Federal Library and Information Network) established by the Federal Library and Information Center Committee (FLICC) of the Library of Congress. Through FEDLINK, FLICC offers all federal agencies cost-effective access to information and library operations support services from commercial sources.

4. SURVEY DESIGN

Target Population

All federal libraries and information centers in the 50 states and the District of Columbia. Foreign branch operations and entities outside of the United States are excluded. For the purposes of this survey, data for Puerto Rico, the Virgin Islands, and U.S. territories are excluded.

To be included in this survey, a library/information center must also meet the following criteria:

(1) be staffed with at least one paid part-time or full-time librarian, technical information specialist, library technician, archivist, or other trained person whose primary function is to assist others in meeting their information needs;

(2) be considered as a federal government operation or receive at least half of its funding from federal sources; and

(3) support the information needs of a federal agency or supply information as part of the agency's mission.

Sample Design

This survey covers the universe of federal libraries and information centers. Major projects involved in developing the survey instrument and defining the universe for the 1994 survey included dissemination of a survey pretest to a sample of 200 facilities in the fall of 1993; the mailing of a locator questionnaire to 3,000 facilities in the spring of 1994 to determine universe eligibility; revision of the survey instrument based on the pretest; and dissemination of a second pretest to a sample of 50 facilities in the fall of 1994.

A variety of sources were searched to develop the initial universe list of approximately 3,200 facilities, which was used as the basis for the locator questionnaire mailing. The primary sources were the Oryx Directory of Federal Libraries and the Federal Library and Information Network (FEDLINK) mailing list. Additional sources included the Federal Health Care Libraries Directory, the U.S. Department of Navy Libraries list, a list of Government Agencies with Public Document Rooms, the Department of Defense (DoD) schools list, the Air Force Library and Information System Address list, and the *U.S. Government Manual*.

The final universe excluded approximately 700 facilities that were overseas (United States Information Service and DoD) and/or elementary and secondary school libraries (DoD and Bureau of Indian Affairs). The overseas facilities were removed because of logistical problems in data collection. The elementary and secondary school libraries under federal agency operation were excluded both to reduce reporting burden and because their mission and function differ from most federal libraries and information centers. NCES includes these schools in a separate survey of School Library Media Centers and Library Media Center Specialists, which is part of the Schools and Staffing Survey (SASS)—see chapter 9. Approximately 1,700 additional facilities were eliminated from the initial universe because they were out of scope of the survey definitions, had combined with another facility, were duplicates of other facilities, or were closed.

Data Collection and Processing

The collection agent for this survey is the U.S. Bureau of the Census. The 1994 survey data were collected and processed between January and September of 1995.

Reference dates. The reporting period for the 1994 survey was the most recent complete fiscal year prior to October 1, 1994. Most data covered the full fiscal year. Data on request and search services were reported for a typical week, defined as a week in which the federal library or information center was open its regular hours (without holidays) and conducted its regular activities. Information reported for the "last 3 years" was reported for the 3 fiscal years from 1992 (ending prior to October 1, 1992) through 1994 (ending prior to October 1, 1994). Information reported for the "next 5 years" was reported for fiscal years from 1995 (ending prior to October 1, 1995) through 1999 (ending prior to October 1, 1999).

Data collection. The 1994 survey was mailed to 1,571 facilities in the United States in January 1995. Of these, 337 were later excluded as out of scope because they did not meet the survey definition of federal libraries and information centers. Thus, there were 1,234 in-scope federal libraries and information centers in the 50 states and District of Columbia.

Only 35 percent of the questionnaires were returned by the March 1995 due date. Rigorous follow-up efforts, including repeated telephone reminders, additional mailings, and special appeals by the FLICC members, were conducted through August. The final response rate was 94.1 percent.

Editing. Prior to keying, the data were manually edited for reporting errors (e.g., when more than one box was marked for items allowing only one answer). The following additional edits were performed after keying: relational edit checks and numeric checks.

Special follow up was required for libraries and information centers which reported reference requests and searches on an annual or other basis instead of weekly. To evaluate the extent of the problem, Census Bureau staff called a sample of cases with possible errors. Approximately 10 percent of the requests and searches data required correction.

Estimation Methods

No adjustment was made for missing information at the unit or item level.

Future Plans

There are no current plans for the next administration of the survey.

5. DATA QUALITY AND COMPARABILITY

Data were not imputed for nonresponse in the 1994 Federal Libraries and Information Centers Survey. *Caution should be exercised when using estimates with item response rates lower than the unit response rate.* Per NCES statistical standards, data are suppressed in published tables if the "total response" (the unit response rate multiplied by the item response rate) is less than 70 percent.

Sampling Error

Because this survey is a universe survey, there is no sampling error.

Nonsampling Error

Coverage error. A comprehensive evaluation of the coverage of the 1994 Federal Libraries and Information Centers Survey revealed some concerns about coverage. Receiving particular consideration was the classification of libraries as out-of-scope, as well as the use of a definition of "federal" library that relied in part on information about the facility's level of federal funding that was provided by the respondent. The study noted that as the 1994 survey's immediate predecessor was conducted more than 15 years earlier, the first task was constructing a survey frame from scratch, a difficult task given that while various directories of federal libraries existed, none of them had the same focus or shared the same definitions as the 1994 survey.

Nonresponse error.

Unit nonresponse. The 1994 survey achieved an overall response rate of 94.1 percent. The response rates by branch of the federal government were as follows:

▸ Judicial Branch	95.2 percent
▸ Legislative Branch	80.0 percent
▸ Executive Branch	
Civilian Departments	75.0–100.0 percent (11 out of 14 were 90 percent or higher)
Military Departments	90.7–96.3 percent
▸ Independent Agencies	90.6–100.0 percent

Item nonresponse. Item response rates in 1994 for published items were as follows: 10 items had a response rate between 92.2 and 94.1 percent. These items primarily consisted of identifying information such as "type of library" and "type of service performed." Another four items had response rates between 86.0 and 89.8 percent. Finally there were three items that obtained response rates of only 76.0-77.5 percent. These items were: size of book print collection (volumes), directional/ready reference requests per typical week, and substantive reference requests per typical week.

Measurement error. Some libraries/information centers reported reference requests and searches on an annual or other basis instead of weekly. A special follow up was conducted by the Census Bureau to evaluate the problem, resulting in correction to about 10 percent of the requests and searches data. *Users should be cautious in their use of these data because only a sample of the lower values was investigated.*

6. CONTACT INFORMATION

For content information the Federal Libraries and Information Centers Survey, contact:

Jeffrey Williams
Phone: (202) 502–7476
E-mail: jeffrey.williams@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Federal Libraries and Information Centers in the United States: 1994, NCES 96–247, by the Governments Division, Bureau of the Census. Washington, DC: 1996.

Data Quality and Comparability

Coverage Evaluation of the 1994 Federal Libraries and Information Centers Survey, NCES 98–269, by J. Curry. Washington, DC: 1998.

Chapter 14: Integrated Postsecondary Education Data System (IPEDS)

1. OVERVIEW

The Integrated Postsecondary Education Data System (IPEDS) is NCES' core postsecondary education data collection program, designed to help NCES meet its mandate to report full and complete statistics on the condition of postsecondary education in the United States. IPEDS is a single, comprehensive system that collects institutional data about all primary providers of postsecondary education. It is built around a series of interrelated surveys designed to collect institution-level data in such areas as enrollments, program completions, faculty, staff, and finances.

Beginning in 1993, survey completion became mandatory for all postsecondary institutions with Program Participation Agreements with the Office of Postsecondary Education, U.S. Department of Education. IPEDS surveys are mandatory for any institution that participates in or is eligible to participate in any federal student financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended (20 USC 1094(a)(17)). For institutions not eligible under Title IV, participation in IPEDS is voluntary. In recent years, these voluntary data were requested only through the Institutional Characteristics survey. Prior to 1993, only national-level estimates from a sample of institutions are available for private less-than-2-year institutions.

In 1998, due to several externally mandated changes and additions to IPEDS, changes in technology for data collection and dissemination, changes in postsecondary education issues, and new expectations for IPEDS, a Redesign Taskforce was charged with recommending changes for the system. The primary recommendation was to switch IPEDS from paper forms to a solely web-based reporting system, which was implemented with the 2000–2001 data collection. IPEDS had been mailing paper forms to institutions on an annual basis since 1986.

It was in 1986 that IPEDS replaced the Higher Education General Information Survey (HEGIS). HEGIS collected data from 1966 to 1986 from a more limited universe of approximately 3,400 institutions accredited at the college level by an association recognized by the Secretary of the U.S. Department of Education. The transition to the IPEDS program expanded the universe to include *all* institutions whose primary purpose is the provision of postsecondary education. The system currently includes about 9,500 postsecondary institutions—including many nonaccredited institutions, as well as schools not accredited at the college level but with vocational/occupational accreditation.

Note that the Office for Civil Rights (OCR) has collaborated with NCES since 1976 regarding the collection of data from postsecondary institutions through Compliance Reports mandated pursuant to Title VI of the Civil Rights Act of 1964, first through HEGIS and then through IPEDS.

SURVEY OF THE UNIVERSE OF POSTSECONDARY INSTITUTIONS

IPEDS collects data annually or biennially through these major components:

- **Institutional Characteristics**
- **Completions**
- **Graduation Rate Survey**
- **Fall Enrollment**
- **Finance**
- **Fall Staff**
- **Faculty Salaries**
- **Institutional Price and Student Financial Aid**

Purpose

To collect institution-level data from all primary providers of postsecondary education—universities and colleges, as well as institutions offering technical and vocational education beyond the high school level.

Components

The IPEDS program consists of several components that obtain information on who provides postsecondary education (institutions), who participates in it and completes it (students), what programs are offered, what programs are completed, and the human and financial resources involved in the provision of institution-based postsecondary education. To avoid duplicative reporting and thus enhance the analytic potential of the database, the various IPEDS data elements and component surveys are interrelated. Several of the surveys used to include different versions of the questionnaire tailored to specific sectors; with the web-based data collection, the tailoring is done through different screens. In general, the data collected from postsecondary institutions granting baccalaureate and higher degrees are the most extensive; the system requests less data from other types of institutions. This feature accommodates the varied operating characteristics, program offerings, and reporting capabilities of postsecondary institutions while yielding comparable statistics for all institutions.

The IPEDS program currently attempts to collects information from approximately 9,500 postsecondary institutions using one or more survey instruments. Because of the requirements for participation in Title IV federal financial aid programs, IPEDS focuses on the 6,600 Title IV institutions. Each of these instruments (or components) is described below; the abbreviation for the survey component is provided after the survey name.

Institutional Characteristics (IC). The core of the IPEDS system is the annual Institutional Characteristics (IC) survey—intended for completion by all currently operating postsecondary institutions in the United States and its outlying areas. As the control file for the entire IPEDS system, IC constitutes the sampling frame for all other NCES surveys of postsecondary institutions. It also helps determine the specific IPEDS screens that are shown to each institution (as it used to determine the specific survey forms that were mailed to each institution). This component collects the basic institutional data that are necessary to sort and analyze not only the IC database, but also all other IPEDS survey databases. The IC survey incorporates many data elements required by state Career Information Delivery Systems (CIDS), thereby reducing or eliminating the need for these organizations to conduct their own surveys.

The number of survey forms used to collect IC data has varied over the years. The 1990–91 IC survey was expanded to incorporate data items previously collected through the IPEDS Institutional Activity (EA) survey, which was phased out in 1989–90; these items now comprise Part D of the Enrollment survey. The version of the survey that a specific institution received used to be a function of its control and program offerings. For the 1999–2000 survey year, which was the last paper collection, there were three versions: IC, IC3, and IC-ADD.

Through 1999, the IC form was mailed to all 4-year, 2-year, and public less-than-2-year institutions; the IC3 form was sent to private less-than-2-year institutions; and the IC-ADD form was sent to all *new* institutions, regardless of control or level. In 1995–96, a short form was developed for use in odd-numbered years to collect minimal data to help maintain the universe and to report on student changes; the full form was used in even-numbered years. Prior to the 1998–99 survey, institutions not eligible for federal financial aid received a different survey form than institutions eligible for federal aid.

IC data are collected for the academic year, which generally extends from September of one calendar year to June of the following year. Specific data elements currently collected for each institution include: institution name, address, telephone number, control or affiliation, calendar system, levels of degrees and awards offered, types of programs, application information, student services, and accreditation. The IC component also collects information on tuition and required fees, room and board charges, books and supplies and other expenses for release on the IPEDS College Opportunities On-Line (IPEDS COOL) web site. These data are made available to prospective students and their parents in order to help them make informed choices about postsecondary education institutions.

Prior to 2000–01, the Institutional Characteristics survey collected instructional activity and unduplicated headcount enrollment for the previous academic year. These data are now collected through the Enrollment (EF) component of IPEDS. The headcount and activity data may be used to compute a standardized, full-time equivalent (FTE) enrollment statistic for the entire academic year. An FTE measure is useful for gauging the size of the educational enterprise at the institution.

Completions (C). This survey collects data annually on recognized degree completions in postsecondary education programs by level (associate's, bachelor's, master's, doctor's, and first-professional) and on other formal awards by length of program. These data are collected by race/ethnicity and sex of recipient and by field of study, which are identified by 6-digit Classification of Instructional Programs (CIP) codes. From 1990 to 1994, racial/ethnic data (by sex and degree/award level) were collected at the 2-digit CIP level. In 1995, there was a major restructuring of the survey to collect race/ethnicity at the 6-digit CIP level and to add additional questions to collect numbers of completers with double majors and numbers of degrees granted at branch campuses in foreign countries. The additional questions were dropped in 2000–01, but a matrix to collect completions data on multiple majors was instituted for optional use in 2001–02 and became mandatory in 2002–03. Completions data on multiple majors will be collected by 6-digit CIP code, award level, race/ethnicity, and sex from those schools that award degrees with multiple majors. (OCR has provided support to collect Completions data since 1976.)

Graduation Rate Survey (GRS). This survey was added in 1997 to help institutions satisfy the requirements of the Student Right-to-Know legislation. The paper version of the annual GRS collected data on the number of students entering an institution as full-time, first-time, degree- or certificate-seeking in a particular year (cohort), by race/ethnicity and sex; length of time to complete; number still persisting; number transferred to other institutions; and number receiving athletically-related student aid and their time to complete. For the 1997–98 GRS, 4-year institutions reported on a 1991 cohort, and less than 4-year institutions reported on a 1994 cohort. The GRS used four different versions to collect data on paper forms. Now that the survey is web-based, institutions see different screens when they are entering data in the web-based data collection system based on a series of screening questions. Also, the number of data items has been reduced. Institutions now provide data on their initial cohort; the number completing within 150 percent of normal time; the number transferred to other institutions; and the number receiving athletically-related student aid. These data allow institutions to disclose and/or report information on the completion or graduation rates and transfer-out rates of these students. Worksheets automatically calculate rates within the web system.

Finance (F). The primary purpose of this annual survey is to collect data to describe the financial condition of postsecondary education in the nation; to monitor changes in postsecondary education finance; and to promote research involving institutional financial resources and expenditures. Specific data elements include current fund revenues by source (e.g., tuition and fees, government, private gifts); current fund expenditures by function (e.g., instruction, research, plant maintenance and operation); physical plant assets and indebtedness; and endowment investments.

Over the years, the various versions of the Finance form have changed. The survey forms for public and private institutions were basically the same except that the public institution form contained three sections with questions pertaining to state and local government financial entities used by the U.S. Bureau of the Census.

The form for private institutions was revised in 1997 to make it easier for respondents to report their financial data according to new standards issued by the Financial Accounting Standards Board (FASB). In an attempt to address reporting issues of proprietary institutions, the form for private institutions was further revised to reflect the General Purpose Financial Statements of these institutions. Again, the reference codes were changed. In addition, with the web-based data collection, the number of data items requested from institutions was greatly reduced in fiscal year (FY) 2000. Due to new accounting standards issued by the Government Accounting Standards Board (GASB), NCES is offering public institutions the option of providing FY 2002 data using a new format that corresponds to the GASB 34/35 standards. This new format, as well as the old version, will be available to institutions as the GASB 34/35 standards are implemented over the next 3 years.

Student Financial Aid (SFA). This component began with a pilot test in 1999, and collected both Institution Price and Student Financial Aid data. The 2000–01 SFA data collection included questions regarding the average amount of financial assistance by type, number of students receiving financial assistance for the previous year, and some contextual items. The tuition and other cost items are now part of the fall Institutional Characteristics (IC) survey; the student financial aid questions are part of the Spring data collection.

Fall Enrollment (EF). This survey collects data annually on the number of full- and part-time students enrolled in postsecondary institutions in the United States and its outlying areas, by level (undergraduate, graduate, first-professional), and by race/ethnicity and sex of student.

Institutions report on students enrolled in courses creditable toward a degree or other formal award; students enrolled in courses that are part of a vocational or occupational program, including those enrolled in off-campus centers; and high school students taking regular college courses for credit. An item that asks for the total number of undergraduates in the entering class (including first-time, transfer, and nondegree students) was added in 2001.

Racial/ethnic data have been collected annually since 1990 (biennially in even-numbered years prior to then). Age distributions are collected in odd-numbered years by student level. Data on state of residence of first-time freshmen (first-time first-year students) and the number that graduated in the past 12 months are collected in even-numbered years (replacing an earlier survey on Residence of First-time Students). Additional questions were asked on students enrolled in branch campuses in foreign countries, those enrolled exclusively in remedial courses, and those enrolled exclusively at extension divisions; however these items are not included in the web-based system. Four-year institutions are also required in even-numbered years to complete enrollment data by level, race/ethnicity, and sex for nine selected fields of study—Education, Engineering, Law, Biological Sciences/Life Sciences, Mathematics, Physical Sciences, Dentistry, Medicine, and Business Management and Administrative Services. Prior to 1996, data were also collected for the fields of Veterinary Medicine and Architecture and Related Programs. The specified fields and their codes are taken directly from *Classification of Instructional Programs (CIP)*. (OCR has supported collection of these data since 1976.)

Fall Enrollment in Occupationally-specific Programs (EP). This survey was incorporated into the IPEDS system in response to the Carl Perkins vocational education legislation. Conducted biennially in odd-numbered years, this survey collected fall enrollment data on students enrolled in occupationally-specific programs at the sub-baccalaureate level, by race/ethnicity and sex of student and by field of study (identified by 6-digit CIP codes). Starting in 1995, total unduplicated counts of students enrolled in these programs were also requested. This survey was discontinued as of the 1999–2000 data collection.

Fall Staff (S). This survey is conducted biennially in odd-numbered years and collects data on the numbers of full- and part-time institutional staff. Specific data elements include: number of full-time faculty by contract length and salary class intervals; number of other persons employed full-time by primary occupational activity and salary class intervals; part-time employees by primary occupational activity; tenure of full-time faculty by academic rank; and new hires by primary occupational activity. Prior to 2001, the survey also requested the number of persons donating (contributing) services or contracted for by the institution. With the exception of contributing/contracted persons, staff data were collected by sex and race/ethnicity.

Between 1987 and 1991, the Fall Staff data were collected in cooperation with the U.S. Equal Employment Opportunity Commission (EEOC). From 1976 through 1991, EEOC collected data on staff through its biennial Higher Education Staff Information (EEO-6) report from all postsecondary institutions within their mandate—that is, institutions that had 15 or more full-time employees. Through the IPEDS program, NCES collected data from all other postsecondary institutions, including all 2- and 4-year higher education institutions with fewer than 15 full-time employees, and a sample of less-than-2-year schools. The 1987–91 IPEDS Fall Staff data files contain combined data from the EEO-6 and the IPEDS staff surveys. Beginning in 1993, all schools formerly surveyed by EEOC were surveyed through the IPEDS Fall Staff survey. (OCR began supporting collection of these data in 1993.)

Employees by Assigned Position (EAP). Beginning with the Winter 2001–02 web-based collection, a new survey, Employees by Assigned Position (EAP), proposed by the National Postsecondary Education Cooperative focus group on faculty and staff, was instituted. This survey was optional the first year and became mandatory in 2002–03. The survey allows institutions to "assign" all faculty and staff to distinct categories. The EAP collects headcount information by full- and part-time status; by function or occupational category; and by faculty and tenure status. Institutions with medical schools are required to report their medical school data separately.

Salaries (SA) (formerly, Salaries, Tenure, and Fringe Benefits of Full-time Instructional Faculty). The primary purpose of this survey was to collect data on the salaries, tenure, and fringe benefits of full-time instructional faculty by contract length, sex, and academic rank; to analyze, from a national perspective, the number and tenure status of faculty members in relation to the number of enrollments and degrees granted for an indication of manpower demand; and to evaluate faculty compensation in relation to institutional financial resources for an indication of the economic status of institutions and of the teaching profession. In previous years, institutions were excluded from the Faculty Salaries survey based on responses to the Institutional Characteristics survey. An institution was excluded if all of its instructional faculty (1) were employed on a part-time basis, (2) were military personnel, (3) contributed their services (e.g., members of a religious order), or (4) taught preclinical or clinical medicine.

Data collected included: total salary outlays (in whole dollars); total number of full-time instructional faculty paid those outlays; number of those faculty who have tenure; who are on tenure track; and who are not on tenure track. These data were collected by rank (professor, associate professor, assistant professor, instructor, lecturer, no academic rank) for men and women on 9/10-month and 11/12-month contracts. Salary outlays, total number of faculty, and tenure status were also collected for full-time faculty on contract schedules other than 9/10 and 11/12 months; however, these data were not collected by rank or sex. Fringe benefits (Part B of the survey) were collected for those full-time instructional faculty reported on Part A. Specific data elements included retirement, tuition, housing and medical dental plans, group life insurance, unemployment and worker's compensation, social security taxes, fringe benefit expenditures (in whole dollars) and the number of full-time faculty covered, by length of contract (9/10 and 11/12-month contracts). This survey was changed from biennial to annual in 1990, and data were not collected in 2000. However, the survey was redesigned for inclusion in the 2001–02 Winter web collection. Much of the same information is currently included except the web survey does not request numbers of faculty by tenure status, but instead collects numbers of faculty by length of contract (less than 9/10 months, 9/10 months, and 11/12 months), rank, sex, and total salary outlay; fringe benefits collection remains the same.

Academic Libraries. First administered in 1966, the Academic Libraries survey was designed to provide concise information on library resources, services, and expenditures for the entire population of academic libraries in the United States. In 1988, the Academic Libraries survey became a part of the IPEDS system and was conducted biennially in even-numbered years. From 1966 to 1988, the Academic Libraries survey was conducted on a 3-year cycle. As of September 2000, this survey ceased to be part of IPEDS. See chapter 11 for a full description of the Academic Libraries Survey.

Consolidated Form (CN and CN-F). A Consolidated Form was used to collect IPEDS data from the institutions eligible for Title IV programs that did not complete the full package of IPEDS surveys described above—that is, those accredited institutions granting only certificates at the sub-baccalaureate level. The Consolidated Form consisted of four or five parts designed to collect, on the same schedule as the regular IPEDS components, minimal data on enrollment (including occupationally-specific programs) and completions by race/ethnicity and sex, as well as data on finance, fall staff, and academic libraries. As of 1996, the "finance" part of the Consolidated Form was on a separate form (CN-F). The purpose and use of the Consolidated data were the same as for the full package of surveys so national data on all accredited institutions could be presented and analyzed. This survey is no longer needed since the web-based data collection system automatically tailors data items for institutions based on selected characteristics and screening questions.

Periodicity

The IPEDS program replaced the HEGIS program in 1986. IPEDS data were collected on paper forms between 1986 and 1999. Since the implementation of the web-based collection of IPEDS data in 2000, most of the surveys are completed by the institutions on an annual basis. However, the survey schedules vary slightly. Institutional Characteristics, Enrollment, Completions, Graduation Rate Survey, Employees by Assigned Position, and Finance are conducted annually. Salaries is an annual survey except for the 2000–01 collection. Fall Staff continues to be collected on a biennial basis in odd-numbered years.

2. USES OF DATA

IPEDS surveys provide a wealth of national-, state-, and institution-level data for analyzing the condition of postsecondary education institutions. For example, the data can be used (with the earlier HEGIS data) to describe long-term trends in higher education. NCES uses IPEDS data in annual reports to Congress on the condition of postsecondary education, statistical digests, profiles of higher education in the states, and other publications. In addition, many requests for information based on IPEDS surveys are received each year from Congress, federal agencies and officials, state agencies and officials, education associations, individual institutions, the media, and the general public. Federal program staff use IPEDS data to address various policy issues. State policymakers use IPEDS data for planning purposes and comparative analysis. Institutional staff use the data for peer analysis.

IPEDS data respond to a wide range of specific educational issues and public concerns. Policymakers and researchers can analyze the types and numbers of postsecondary institutions; the number of students, graduates, first-time freshmen, and graduate and professional students by race/ethnicity and sex; the status of postsecondary vocational education programs; the number of individuals trained in certain occupational and vocational fields by race/ethnicity, sex, and level; the resources generated by postsecondary institutions; patterns of expenditures and revenues of institutions; changes in tuition and fees charged; completions by type of program, level of award, race/ethnicity, and sex; faculty salaries and composition; and many other topics of interest.

The IPEDS universe also provides the institutional sampling frame used in all NCES postsecondary surveys, such as the National Postsecondary Student Aid Study (NPSAS) and the National Study of Postsecondary Faculty (NSOPF). Each of these surveys uses the IPEDS institutional universe for its first-stage sample and relies on IPEDS survey results on enrollment, completions, or staff to weight its second-stage sample.

OCR supports the collection of IPEDS enrollment, completions, and fall staff data, and uses these data to produce such reports as *2001 U.S. Accredited Postsecondary Minority Institutions*.

3. KEY CONCEPTS

Described below are several key concepts relevant to the IPEDS program. For additional terms, refer to the *IPEDS Glossary* (NCES 95–822).

Postsecondary Education. The provision of a formal instructional program whose curriculum is designed primarily for students who are beyond the compulsory age for high school. Programs whose purpose is academic, vocational, or continuing professional education are included. Excluded are avocational and adult basic education programs.

Institution of Higher Education (IHE). Prior to 1996, an IHE was defined as an institution accredited at the *college* level by an accrediting agency or association recognized by the Secretary of the U.S. Department of Education—and indicated as such in the database by the presence of a Federal Interagency Committee on Education (FICE) code. IHEs were legally authorized to offer at least a 1-year program of study creditable toward a degree.

Degree-granting Institution. Any institution offering an associate's, bachelor's, master's, doctor's, or first-professional degree. Institutions that grant only certificates or awards of any length (less than 2 years, or 2 years or more) are categorized as nondegree-granting institutions.

Branch Institution. A campus or site of an educational institution that is not temporary, that is located in a community beyond a reasonable commuting distance from its parent institution, and where organized *programs* of study (not just courses) are offered. This last criterion is the most important. It means that at least one degree or award program can be completed entirely at the site without requiring any attendance at the main campus or any other institution within the system.

OPEID Code. An 8-digit identification code developed by the U.S. Department of Education's Office of Postsecondary Education (OPE) for the Postsecondary Education Participants System (PEPS). Presence of a valid OPEID in the database indicates that the school has a Program Participation Agreement with the Department and is currently eligible to participate in Title IV federal financial aid programs (e.g., Pell Grants, Stafford Loans, College Work-study). The first 6 digits of the OPEID are the old FICE code and represent the ID of the institution. The last 2 digits identify the various campuses or additional locations. For the main campus, the last 2 digits will always be "00." If the last 2 digits are numeric (e.g., 01, 02, 03), the institution is a branch campus or other location of an eligible main campus and is listed separately in PEPS. If the last 2 digits of the OPEID are of the form A1, A2, etc., the entity is separately identified in IPEDS for reporting purposes.

Occupationally-specific Program. An instructional program below the bachelor's level, designed to prepare individuals with entry-level skills and training required for employment in a specific trade, occupation, or profession related to the field of study.

CIP Code. A 6-digit code, in the form xx.xxxx, that identifies instructional program specialties within educational institutions. The codes are from the NCES publication, *A Classification of Instructional Programs (CIP)*.

4. SURVEY DESIGN

Target Population

All institutions (in the 50 states, the District of Columbia, and outlying areas) whose primary purpose is the provision of postsecondary education. The IPEDS universe includes all institutions and branches that offer a full program of study (not just courses); freestanding medical schools, as well as schools of nursing, schools of radiology, etc., within hospitals; and schools offering occupational and vocational training with the intent of preparing students for work (e.g., a modeling school that trains for professional modeling, but not a charm school).

The IPEDS universe of postsecondary institutions does *not* include institutions that are not open to the general public (training sites at prisons, military installations, corporations); hospitals that offer only internships or residency programs, or hospitals that offer only training as part of a medical school program at an institution of higher education; organizational entities providing only noncredit continuing education; schools whose only purpose is to prepare students to take a particular test, such as the CPA or Bar exams; and branch campuses of U.S. institutions in foreign countries. Relevant data from such locations or training sites are to be incorporated into the data reported by the main campus or any other institution or branch campus in the system that is most appropriate.

Eligibility for Title IV federal financial aid, while not a requirement for inclusion in the universe, defines a major subset of all postsecondary institutions. Prior to 1996, aid-eligible institutions were self-identified as IHEs or were identified as aid-eligible from responses to items on the Institutional Characteristics survey. Beginning in 1996, the subset of aid-eligible institutions is validated by matching the IPEDS universe with the PEPS file maintained by OPE. OPE grants eligibility to institutions to participate in Title IV federal financial aid programs.

In establishing the PEPS file, the U.S. Department of Education discontinued its tradition of distinguishing institutions accredited at the college level from institutions accredited at the occupational/vocational level. Therefore, it is no longer possible for NCES to maintain a subset of accredited institutions at the college level (IHEs). Beginning with the 1997 IPEDS mailout and on the 1996 and subsequent data files, institutions are classified by whether or not they are eligible to participate in Title IV financial aid programs and whether or not they grant degrees (as opposed to awarding only certificates).

Sample Design

Prior to 1993, data were collected from a representative sample of about 15 percent of the universe of private, for-profit, less-than-2-year institutions. However, the Higher Education Act of 1992 mandated the completion of IPEDS surveys for all institutions that participate or are applicants for participation in any federal student financial assistance program authorized by Title IV of the Higher Education Act of 1965, as amended. Thus, beginning with the 1993 IPEDS mailout, NCES surveys in detail *all* postsecondary institutions meeting this mandate.

Data Collection and Processing

The U.S. Bureau of the Census served as the data collection agent for the IPEDS surveys from 1990 through the 1999–2000 survey. Survey forms were either submitted directly to the Census Bureau by the institutions or through a central or state coordinating office. The web-based system was implemented with the 2000–01 survey, with different contractors developing the web site and managing the collection process.

The IPEDS institution-level data collection allows for aggregation of results at various levels and permits significant controls on data quality through editing. Attempts are made to minimize institutional respondent burden by coordinating data collection with the states and with other offices and agencies that regularly collect data from institutions.

Reference dates. Data for the IPEDS surveys are collected for a particular school year, term, or fiscal year, as follows:

- The Institutional Characteristics (IC) survey collects data for the entire academic year, generally starting in September or with the fall term if there is one. For example, data collected in 2002 pertain to the 2002–03 academic year, usually September 2002 through June 2003. In the case of schools operating on a 12-month calendar, the collection period runs from September 2002 through August 2003.
- The Completions survey collects data for an entire 12-month period, which is defined as July 1 through June 30; in some instances, start dates may vary slightly by institution.
- For the Graduation Rate Survey, the majority of institutions report on the status of students in their cohort (either a fall cohort or a full-year cohort) as of August 31. Section V requests data on students enrolled during the period September 1 through August 31 of the year prior to submission of the report.

- The Finance survey collects data for the institution's most recent fiscal year, generally ending before October 1 (although some institutions may have other ending dates). Thus, data collected in spring 2003 pertain to the fiscal year just ended, FY 2002.

- The Student Financial Aid survey collects the average amount of financial assistance and the number of students receiving financial assistance for the prior academic year.

- The Fall Enrollment survey (and previously the Fall Enrollment in Occupationally-specific Programs survey) collects data for a single point in time during the fall term, usually recorded as of the institution's official fall reporting date or October 15. If there is no fall term or class activity, institutions are asked to report zero enrollment. Part D of the survey now collects unduplicated headcount and instructional activity (formerly part of IC); these data are reported for the 12-month period that ended prior to September 1 of the reporting year.

- The Fall Staff survey collects data on employees who were on the payroll of the institution as of November 1 of the survey year and data on new hires from July 1 through October 30 of the survey year. Prior to the 2001 collection, institutions reported as of October 1.

- The Salaries survey (formerly Salaries of Full-time Instructional Faculty) collects data on the number of full-time instructional faculty as of November 1 (formerly October 1) of the survey year. Salaries and fringe benefits reflect the full academic year (e.g., academic year 2002–03, with data reported in winter 2002).

- The Student Financial Aid survey collects financial aid information (for the prior academic year) in the spring collection.

Data collection. Since institutions are the primary unit of data collection, institutional units must be defined as consistently as possible. The IPEDS program does not request separate reports from more than one component within an individual institution; however, separate branch campuses are asked to report as individual units. Following the HEGIS model, the IPEDS program is intended to collect data from each institution in a multi-institutional system and each separate branch in a multi-campus system.

Between 1993 and 1996, NCES began to examine the universe of accredited institutions in order to form a crosswalk between the IPEDS data files and those maintained by OPE for student financial aid purposes. During this period, OPE discontinued its policy of differentiating institutions by level of accreditation—that is, those accredited at the college level (formerly the HEGIS universe) versus those with occupational/vocational accreditation. Since the IPEDS system could no longer identify institutions with college-level accreditation, a new approach was developed to categorize institutions for mailout and analysis purposes. Beginning with the 1997 mailout, the IPEDS universe was subdivided according to: (1) accreditation status, (2) level of institution, and (3) degree-granting status. The current web-based system considers Title IV status rather than accreditation.

Prior to the development of the web-based data collection system, IPEDS survey forms were mailed to institutions based upon the information provided on the prior year's Institutional Characteristics survey—control and highest level of offering (which determined an institution's sector) combined with accreditation status. Institutions that were not accredited, and thus not eligible for federal student financial aid, were asked to complete only the Institutional Characteristics survey. All accredited institutions that either (1) grant an associate's or higher degree or (2) offer a certificate program above the baccalaureate level received a full packet of surveys—Institutional Characteristics; Completions; Fall Enrollment; Fall Enrollment in Occupationally-specific Programs; Fall Staff; Finance; Graduation Rate Survey; Salaries of Full-time Instructional Faculty; and Academic Libraries. All other accredited institutions (i.e., those granting only certificates at the sub-baccalaureate level) were required to complete the Institutional Characteristics survey, the Graduation Rate Survey (if applicable), and a Consolidated Form.

Institutions not in the IPEDS universe, but identified as "possible adds," received an IC-ADD survey. With the web system, these same "new" schools enter similar data directly into the system. Schools targeted as "possible adds" are identified from many sources, including a universe review done by state coordinators, a review of the PEPS data file from OPE, and information received from the institutions themselves. Institutions are added to the universe if they respond that their primary mission is the provision of postsecondary education as defined in the survey.

Prior to 2000–01, most of the data collection from the institutions that completed the full complement of IPEDS surveys was done through state-level higher education agencies. Coordinators were given the option of assisting NCES in various ways, including mailing packages to schools, coordinating nonresponse follow up, mailing survey forms back to NCES, resolving errors, and maintaining the universe. Beginning in 2000–01, an electronic

BEST COPY AVAILABLE

coordination system (or tree) is used to route institutional and/or state responses, as applicable, through the state coordinators. Coordinators may continue to choose the sectors and institutions they wish to monitor (e.g., they can identify "just 4-year schools" or almost specify on a one-by-one basis; coordinators can also still choose to "view" the data only, or actually review, approve, and "lock" the data). In many states, IPEDS institutional data are provided by the state higher education agency from data collected on state surveys. Alternatively, state agencies may extract data from IPEDS rather than conduct their own surveys.

To ease respondent burden, the Institutional Characteristics web screens include previously reported data, and survey respondents are instructed to update the previous data if necessary and to provide current information for items such as tuition and required fees, and room and board charges. (In earlier years, IC forms were preprinted with prior-year survey responses for those items that generally were not expected to change from year to year.) Questionnaires/screens for other IPEDS surveys contain selected preprinted information, such as CIP codes and program titles on the Completions and Enrollment surveys.

Prior to the Fall 2000 survey, institutions reported IPEDS data by mail on paper forms or diskettes, by fax, or electronically through the Internet. Two methods were available: the first method involved a predetermined ASCII record layout, available for all surveys except Institutional Characteristics. For Fall Enrollment and the Graduation Rate Survey, downloadable software was also available, allowing for data entry as well as preliminary editing of the data before transmission to the Census Bureau.

Mailouts of all applicable surveys took place in July of the survey year, except in 1998–99 when forms were not mailed until August. Due dates varied by survey. Extensive follow-up for survey nonresponse was conducted during the 6 months following each survey's due date. Initially, reminder letters were mailed, encouraging nonresponding institutions to complete and return their forms. Subsequently, the Postsecondary Education Telephone System (PETS) was used to collect critical data by telephone from representatives of institutions for which IPEDS state coordinators are not responsible for follow-up. With the web system, institutions receive letters in mid-July containing IDs and passwords and instructions for registering their keyholders. Follow up is conducted either with the Chief Executive Officers (CEOs) if there is no registered keyholder, or directly with the keyholder.

Institutions found to be out-of-scope during data collection are deleted from the universe. These deletions result from formal notification from IPEDS state coordinators and follow-up telephone calls. Included in the deletions are: (1) duplicates of other institutions on the file; (2) institutions that closed or merged with another institution, and thus are no longer legitimate institutions or branches; (3) institutions that no longer offer postsecondary programs; and (4) schools that do not conform to the IPEDS definition of an institution or branch. The final IPEDS universe is also adjusted to reflect institutions that changed from one sector to another.

The following collection schedule was planned for the 2002–2003 academic year:

- Fall 2002—The Fall 2002 collection (September 9–November 5, 2002) included the Institutional Characteristics and Completions components.
- Winter 2002–03—The Winter 2002–03 collection (November 25, 2002–February 5, 2003) included Employees by Assigned Position, Salaries, Fall Staff (optional) and Enrollment. (Institutions may complete the Enrollment component in either winter or sprint).
- Spring 2003—The Spring 2003 collection (March 5–April 30, 2003) included the collection of Enrollment (both fall and full year), Finance, Student Financial Aid information, and Graduation Rates data.

The current IPEDS universe includes approximately 9,600 postsecondary institutions and 80 administrative units.

Editing. IPEDS data are edited for reporting and processing errors. All data, whether received on paper forms, diskettes, electronically through the Internet, or through the PETS system, went through the same editing process to verify internal and inter-year consistency. Addition checks were performed by adding down or across columns and comparing generated totals with reported totals. If the reported total differed from the generated total but was within a designated range, the reported total was replaced by the generated total and the cell was flagged with the proper imputation code. Otherwise, institutions were contacted to resolve the discrepancies. Data collected on the web surveys are edited in a similar fashion except that the web system automatically generates all totals. In addition, all errors must be resolved prior to "locking" by the institution.

All program entries (submissions by field) on the Completions and Institutional Characteristics components are checked for CIP code validity against *A Classification of*

Instructional Programs. When possible, missing data items are identified during the edit process; formerly, they were resolved during telephone follow-up with institutions. Imputation is performed when certain key data items are not reported. For total nonresponding institutions, data are also imputed. Final quality control procedures are performed when all institutions have responded or been imputed. (See Estimation Methods below for the imputation methods used.)

Data also are compared between IPEDS survey components. For instance, if a change in award level on the Institutional Characteristics survey triggers a sector change, it is verified against the Completions survey or the Enrollment survey. All award levels and first-professional programs listed on the Institutional Characteristics survey are checked against the Completions survey. Checks are made to ensure the cohort reported on the Graduation Rate Survey is comparable to the data reported on the Fall Enrollment survey for the appropriate cohort year. Large discrepancies are flagged and all errors must be resolved before keyholders can lock their data. Data are also checked for consistency with prior-year responses (if available). If the differences are sufficiently large to trigger an edit flag, institutions must confirm or explain the discrepancy.

Estimation Methods

Imputation is done to compensate for nonresponding institutions—both total nonresponse and partial nonresponse to specific data items. Prior to 1993, all sectors were surveyed and a sample of private less-than-2-year institutions was conducted to obtain national estimates for fall enrollment, completions, finance and fall staff; these data were weighted and subject to sampling error. Starting in 1993, the IPEDS eliminated the sample of the private less-than-2-year institutions and continue to survey the entire universe of postsecondary institutions; therefore, no weighting is conducted.

Imputation. The IPEDS system used cold-deck (updated by ratio methods to reflect the change) and hot-deck imputation procedures to adjust for partial or total nonresponse to a specific survey instrument. Current imputation for missing data is performed after all editing is completed. IPEDS uses several methods of imputation depending on the availability of prior year data including a "carry forward" method, group means, and "nearest neighbor." All IPEDS surveys use the same imputation flags. Institutions that are entirely imputed may be identified on the file by their response status and imputation type codes. For responding institutions that are edited or partially imputed, the affected items may be identified by the associated item imputation flags.

Recent Changes

Key changes to the IPEDS program in the 1990s are summarized below:

- Beginning in 1995–96, Part D of the IC form includes questions about tuition previously asked in other IC form types. Institutions were asked their method(s) of charging tuition and, from that response, were directed toward the appropriate set of follow-up questions. Institutions that charge tuition both by program (for vocational/occupational programs) and by semester or term (for academic programs) were requested to report both methods in different questions. If the institution charges tuition by only one of the methods, it reports the amount charged in the appropriate question. Prior to 1995–96, different IC forms were used for program versus semester/term charges.
- The IPEDS program no longer differentiates between accredited college-level institutions and postsecondary institutions with occupational or vocational accreditation. Beginning with the 1997 IPEDS mailout and on the 1996 and subsequent data files, institutions are classified by whether or not they are eligible to participate in Title IV financial aid programs and whether or not they grant degrees, not by highest level of offering.
- As of 1996 in the Fall Enrollment survey, 4-year institutions are no longer required to report enrollment data by level, race/ethnicity, and sex for the fields of Veterinary Medicine and Architecture and Related Programs.
- In 1997, GRS was added to the IPEDS program to help institutions satisfy the requirements of the Student Right-to-Know legislation.
- Beginning with the 1998–99 Institutional Characteristics survey, data on credit and contact hour activity for the 12-month period and the fall term and data on the unduplicated count of students by level for the 12-month period are collected from all but new postsecondary institutions. In earlier years, data on credit and contact hour activity were collected only from institutions eligible for federal financial aid. Also, items on summer session and extension division activity were dropped from the 1998–99 IC survey.
- NCES added several new items for the 1999–2000 Institutional Characteristics survey.
- In 1999, NCES collected selected data items in a pilot test through a web-based survey: tuition and fees for entering students, room and board, books and supplies, and information on students receiving financial aid. These items have been incorporated, where appropriate, in the

BEST COPY AVAILABLE

redesigned IPEDS data collection, implemented in 2000–01.

- In 2000–01, NCES converted IPEDS to a totally web-based data collection system. The content of the survey "forms" was revised and reduced in scope and the procedures for collecting data vary considerably from those used in prior years. The first year, two collection cycles were implemented: Fall 2000 collected IC and Completions data and Spring 2001 included the Enrollment, Student Financial Aid, Finance, and Graduation Rates components. Subsequent years include a Winter cycle to collect Employees by Assigned Position, Salaries, and Fall Staff data.

Future Plans

IPEDS plans to continue with three separate data collections (fall, winter, and spring) in future years. Data items may be modified to better reflect current issues in postsecondary education as recommended by the IPEDS Technical Review Panel (TRP).

5. DATA QUALITY AND COMPARABILITY

Data element definitions have been formulated and tested to be relevant to all providers of postsecondary education and consistent among components of the system. A set of data elements has been established to identify characteristics common to all providers of postsecondary education, and specific data elements have been established to define unique characteristics of different types of providers. Interrelationships among various components of IPEDS have been formed to avoid duplicative reporting and to enhance the policy relevance and analytic potential of the data. Through the use of "clarifying" questions that ask what was or was not included in a reported count or total or the use of caveats that supplement the web collection, it is possible to address problems in making interstate and interinstitutional comparisons. Finally, specialized, but compatible, reporting formats have been developed for the different sectors of postsecondary education providers. This design feature accommodates the varied operating characteristics, program offerings, and reporting capabilities that differentiate postsecondary institutional sectors, while yielding comparable statistics for some common parameters of all sectors.

Sampling Error

Only the data collected prior to 1993 from a sample of private less-than-2-year institutions are subject to sampling error. With this one exception, the HEGIS and IPEDS programs include the universe of postsecondary institutions.

Nonsampling Error

IPEDS data are subject to such nonsampling errors as errors of design, reporting, processing, nonresponse, and imputation. To the extent possible, these errors are kept to a minimum by methods built into the survey procedures.

The sources of nonsampling error in IPEDS data vary with the survey instrument. In the Fall Enrollment survey, major sources of nonsampling error are classification problems, unavailability of needed data, misinterpretation of definitions, and operational errors. Possible sources of nonsampling error in the Finance survey include nonresponse, imputation, and misclassification. The primary sources of nonsampling error in the Completions survey are differences between the NCES program taxonomy and taxonomies used by colleges, classification of double majors and double degrees, operational problems, and survey timing.

Coverage error. Coverage error in the IPEDS system is believed to be minimal. For institutions that are eligible for Title IV federal financial aid programs, coverage is almost 100 percent. Schools targeted as "possible adds" are identified from many sources, including a universe review done by state coordinators, a review of the PEPS file from OPE, and the institutions themselves.

Nonresponse error. Since 1993, all institutions entering into Program Participation Agreements (PPAs) with the U.S. Department of Education are required by law to complete the IPEDS package of surveys. Therefore, overall unit and item response rates are quite high for all surveys for these institutions. Data collection procedures, including extensive mail and telephone follow-ups, also contribute to the high response rates. Imputation is performed to adjust for both partial and total nonresponse to a survey. Because response rates are so high, error due to imputation is considered small.

Unit nonresponse. Overall unit response rates are high for all surveys. For example, the percent of all institutions that responded to various IPEDS surveys are listed below:

1996–97 Institutional Characteristics	92.0
1996–97 Faculty Salaries	92.9
1996 Fall Enrollment	95.0

1995–96 Completions	94.5
1995 Fall Staff	86.9
FY 95 Finance	82.6

Since the implementation of the web collection, Title IV institutional response rates range from about 89 percent on the SFA survey to about 98 percent on IC. (See chapter 11 for response rates for the Academic Libraries Survey.)

By sector, the response rates are highest for public 4-year or higher institutions and lowest for private for-profit institutions, especially the less-than-2-year institutions. The 1994 Academic Libraries and the FY 95 Finance public use data files are limited to IHEs because the response rate for postsecondary institutions not accredited at the collegiate level was quite low (74.1 percent in the Finance survey and less than 50 percent in the Academic Libraries survey).

Item nonresponse. Most participating institutions provide complete responses on all items. Telephone follow up is used to obtain critical missing items. For the Fall Staff data, partial nonresponse is relatively rare.

Measurement error. NCES strives to minimize measurement error in IPEDS data by using various quality control and editing procedures. New questionnaire forms or items are field tested and/or reviewed by experts prior to use. To minimize reporting errors in the Finance survey, NCES uses national standards for reporting finance statistics. Wherever possible, definitions and formats in the Finance survey are consistent with those in the following publications: *College and University Business Administration. Administrative Services, Financial Accounting and Reporting Manual for Higher Education; Audits of Colleges and Universities*, and *HEGIS Financial Reporting Guide*.

The classification of students appears to be the main source of error in the Enrollment survey. Institutions have had problems in correctly classifying first-time freshmen, other first-time students, and unclassified students for both full-time and part-time categories. These problems occur most often at 2-year institutions (both public and private) and private 4-year institutions. In the 1977–78 HEGIS validation studies, misclassification led to an estimated overcount of 11,000 full-time students and an undercount of 19,000 part-time students. Although the ratio of error to the grand total was quite small (less than 1 percent), the percentage of errors was as high as 5 percent for detail student levels and even higher at certain aggregation levels. (See also Data Comparability below.)

Data Comparability

The definitions and instructions for compiling IPEDS data have been designed to minimize comparability problems. However, survey changes necessarily occur over the years, resulting in some issues of comparability. Also, postsecondary education institutions vary widely, and hence, comparisons of data provided by individual institutions may be misleading. Specific issues related to the comparability of IPEDS data are described below.

Imputation. Imputed data are on file for institutions with partial or total nonresponse. *Caution should be exercised when comparing institutions for which data have been imputed since these data are intended for computing national totals and not intended to be an accurate portrayal of an institution's data. Users should also be cautious when making year-to-year enrollment comparisons by state.* In some cases, state enrollment counts vary between years as a result of imputation rather than actual changes in the reported enrollment data. To avoid misinterpretation, users should always check the response status codes of individual institutions to determine if a large proportion of data was imputed.

Classification of institutions. Beginning in 1996, the subset of IPEDS institutions eligible to participate in Title IV federal financial student aid has been validated by matching the IPEDS universe with the PEPS file maintained by OPE. Previously, institutions were self-identified as aid-eligible from the list of IHEs and responses to the Institutional Characteristics survey.

Another note of caution concerns the use of form type (e.g., EF1, EF2, or CN) versus institutional sector. Forms were mailed to institutions based on information provided on the *prior* year's IC survey. When schools returned forms that were inconsistent with the sector in which they were identified on the earlier IC survey, NCES attempted to determine their proper sector. Then, either the school's sector was adjusted or the data returned were adjusted to conform to the proper survey form. Even if the institution's characteristics change in the current IC survey, completions can properly be reported for the prior sector. However, the completions from any new programs will only be reported in subsequent years. For these reasons, *it is important to query the counts of completions for the degree levels needed rather than the sector; otherwise, legitimate completions will be missed in calculations or the* *number of schools identified for a specified highest offering (e.g., baccalaureate) may be over- or understated.*

Fields of study. In analyzing Completions data by field of study, users must remember that the data represent programs, not schools, colleges, or divisions within institutions. For example, some institutions might have a few computer and information science programs organized and taught within a business school. However, for IPEDS reporting purposes, the degrees are classified and counted within the computer and information science discipline division.

Reporting periods. The data collected through IPEDS surveys for any one year represent two distinct time periods. The Institutional Characteristics, Enrollment (most parts), Fall Staff, and Salaries, and Employees by Assigned Position data represent an institution at one point in time, the fall of the school year; whereas, the Instructional Acitivy portion of the Enrollment survey, Student Financial Aid, Finance, and Completions data cover an entire 12-month period or fiscal year. For some indicators, fall data are used in conjunction with 12-month data in NCES reports, and readers should be cognizant of the differences in time periods represented.

Questionnaire changes. Over the years, the IPEDS survey forms have undergone revisions, which may have an impact on data comparability. Users should consider the following:

- The number of forms used to collect IC data has varied between survey administrations. However, form type is recoded in the IC data file to maintain prior types.
- As of the 1994–95 academic year, the Completions survey is substantially different from earlier surveys. The basic changes are: (1) there is only one survey form, collecting counts of degrees and other awards at all levels; (2) race/ethnicity data are collected by award level for detailed fields of study; and (3) data are/were collected in two clarifying questions to determine the extent of double majors and awards conferred at branch campuses in foreign countries.
- Beginning in 1995–96, institutions that charge tuition both by program and by semester or term report the amounts for each method in different questions on the same form. If the institution uses only one method, it reports the amount charged in the appropriate question. Prior to 1995–96, different IC forms were used for program versus semester/term charges. (Beginning in 1999–2000, the IC survey will request separate reporting of tuition, required fees, and per-credit-hour charge for in-district, in-state, and out-of-state students.)
- Beginning in fall 1995, the salary class intervals were revised for the Fall Staff survey; this may affect historical comparisons and analysis. In addition, a new Part C, "All Other Full-time Employees," was added to the Fall Staff survey.
- To enhance the comparability and utility of the finance data, NCES has made several improvements in the reporting of IPEDS financial statistics: (1) information is requested on expenditures by object (salaries, employee benefits, library acquisitions, and utilities); (2) a series of clarifying questions determine what is included/excluded from reported current fund expenditures; (3) a section is included on expenditures for student scholarships and fellowships from federal, state, local, and institutional sources; and (4) appropriations for hospitals are separated from appropriations for the educational institution.
- The Finance F1-A form for private institutions was revised in 1997 to make it easier for respondents to report their financial data according to the new standards issued by the Financial Accounting Standards Board. In an attempt to address reporting issues of proprietary institutions, the F1-A was revised in 1999 to reflect the financial statements of these institutions. This split the F1-A into two forms: F2 for private, not-for-profit institutions and F3 for private for-profit institutions.

Comparisons with HEGIS. *Caution must be exercised in making cross-year comparisons of institutional data collected in the IPEDS system with data collected in the HEGIS system.* The IPEDS surveys request separate reporting by all institutions and their branches as long as each entity offers at least one complete program of study. Under the HEGIS program, only separately accredited branches of an institution were surveyed as separate entities; branches that were *not* separately accredited were combined with the appropriate entity for purposes of data collection and reporting. Therefore, an institution may have several entities in the IPEDS system where only one existed in the HEGIS system.

Comparison with the Survey of Earned Doctorates. Like the IPEDS Completions survey, the Survey of Earned Doctorates (SED, see chapter 19) also collects data on doctoral degrees, but the information is provided by doctorate recipients rather than by institutions. The number of doctorates reported in the Completions survey is slightly higher than in SED. This difference is largely attributable to the inclusion of nonresearch doctorates (primarily in theology and education) in the Completions survey. The discrepancies in counts have been generally consistent since 1960, with ratios of IPEDS-to-SED counts ranging from 1.01 to 1.06. Differences in the number of doctorates within a given field may be greater than the overall difference because a respondent to SED may classify his/her specialty differently than the institution reports the field in the Completions survey.

6. CONTACT INFORMATION

For content information on the IPEDS system, contact:

Susan G. Broyles
Phone: (202) 502–7318
E-mail: susan.broyles@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Basic Statistics from Non-Collegiate Institutions, 1990, NCES 92–053, by S.G. Broyles. Washington, DC: 1992.

IPEDS Graduation Rate Survey: Guidelines for Survey Respondents, NCES 98–904, by S. Broyles. Washington, DC: 1998.

IPEDS Manual for Users. Washington, DC: 1994.

IPEDS Training Manual #1, NCES 93–195, by S.G. Broyles. Washington, DC: 1992.

IPEDS Training Manual #2, NCES 93–196, by S.G. Broyles. Washington, DC: 1992.

Uses of Data

Classification of Instructional Programs, 1990 Update, NCES 91–396, by R. Morgan and W. Freund. Washington, DC: 1991.

Integrated Postsecondary Education Data System Glossary, NCES 95–822, by S. Broyles. Washington, DC: 1995.

Chapter 15: National Study of Postsecondary Faculty (NSOPF)

1. OVERVIEW

The National Study of Postsecondary Faculty (NSOPF) is conducted to provide information on postsecondary faculty and instructional staff: their academic and professional background, sociodemographic characteristics, and employment characteristics such as institutional responsibilities and workload, job satisfaction, and compensation. Thus far, there have been three NSOPF administrations—one in the 1987–88 academic year, a second one in the 1992–93 academic year, and the third one in the 1998–99 academic year. The first cycle was conducted with a sample of institutions, faculty, and department chairpersons. The second and third cycles were limited to surveys of institutions and faculty, but with a substantially expanded sample of public and private, not-for-profit institutions and faculty.

PERIODIC SURVEY OF A SAMPLE OF POSTSECONDARY INSTITUTIONS AND THEIR FACULTY

NSOPF includes:

- Institution Survey
- Faculty Survey
- Department Chairperson Survey (1987–88 only)

Purpose

To provide a national profile of postsecondary faculty: their professional backgrounds, responsibilities, workloads, salaries, benefits, and attitudes.

Components

NSOPF consists of two surveys, one for institutions and the other for faculty. Institutions receive both an *Institution Survey* and a request to provide a faculty list. The *Faculty Survey* is sent to faculty and other instructional staff sampled from the lists provided by the institutions. The 1987–88 NSOPF also included a *Department Chairperson Survey*.

Institution Survey. The Institution Survey obtains information on: the numbers of full- and part-time instructional and noninstructional faculty, as well as instructional personnel without faculty status; tenure status of faculty members (based on definitions provided by the institution); institution tenure policies and changes in policies on granting tenure to faculty members; the impact of tenure policies on the influx of new faculty and on career development; the growth and promotion potential for existing nontenured junior faculty; the benefits and retirement plans available to faculty; and the turnover rates of faculty at the institution. The survey is completed by an institutional respondent designated by the Chief Administrative Officer (CAO) at each sampled institution.

Faculty Survey. This survey addresses the following issues as they relate to postsecondary faculty: background characteristics and academic credentials; workloads and time allocation between classroom instruction and other activities such as research, course preparation, consulting, public service, doctoral or student advising, conferences, and curriculum development; compensation and the importance of other sources of income such as consulting fees, royalties, etc., or income-in-kind; roles and differences, if any, between full- and part-time faculty in their participation in institutional policymaking

BEST COPY AVAILABLE

and planning; faculty attitudes toward their jobs, their institutions, higher education, and student achievement in general; changes in teaching methods and the impact of new technologies on teaching techniques; career and retirement plans; differences between individuals who have instructional responsibilities and those who have no instructional responsibilities (e.g., those engaged only in research); and differences between those with teaching responsibilities but no faculty status and those with teaching responsibilities and faculty status. Eligible respondents for this survey are faculty members sampled from lists provided by institutions involved in the study. These lists are compiled by the Institutional Coordinator designated by the CAO at each sampled institution.

Department Chairperson Survey. Conducted only in 1987–88, this survey collected information from over 3,000 department chairpersons on faculty composition in departments, tenure status of faculty, faculty hires and departures, hiring practices, activities used to assess faculty performance, and professional and developmental activities.

Periodicity

The NSOPF was conducted in 1987–88, 1992–93, and 1998–99. The next round is planned for 2003–04.

2. USES OF DATA

NSOPF provides valuable data on postsecondary faculty that can be applied to policy and research issues of importance to federal policymakers, education researchers, and postsecondary institutions across the United States. For example, NSOPF data can be used to analyze whether the postsecondary labor force is declining or increasing. NSOPF data can also be used to analyze faculty job satisfaction and how it correlates with an area of specialization, and also how background and specialization skills relate to present assignments. Comparisons can be made on academic rank and outside employment. Benefits and compensation can be studied across institutions, and faculty can be aggregated by sociodemographic characteristics. Because NSOPF is conducted periodically, it also supports comparisons of data longitudinally.

The Institution Questionnaire includes items about:

- the number of full- and part-time faculty (i.e. instructional and noninstructional), as well as instructional personnel without faculty status, and their distributions by employment (i.e. full-time, part-time) and tenure status (based on the definitions provided by the institution);
- institutional tenure policies and changes in policies on granting tenure to faculty members;
- the impact of tenure policies on the number of new faculty and on career development;
- the growth and promotion potential for existing nontenured junior faculty;
- the procedures used to assess the teaching performance of faculty and instructional staff;
- the benefits and retirement plans available to faculty; and
- the turnover rates of faculty at the institution.

The Faculty Questionnaire addresses such issues as respondents' employment, academic and professional background, institutional responsibilities and workload, job satisfaction, compensation, sociodemographic characteristics, and opinions. The questionnaire is designed to emphasize behavioral rather than attitudinal questions in order to collect data on who the faculty are, what they do, and whether, how and why the composition of the nation's faculty is changing. The Faculty Questionnaire includes items about:

- background characteristics and academic credentials;
- workloads and time allocation between classroom instruction and other activities such as research, course preparation, consulting, work at other institutions, public service, doctoral or student advising, conferences, and curriculum development;
- compensation and the importance of other sources of income, such as consulting fees, royalties, etc. or income-in-kind;
- the number of years spent in academia, and the number of years with instructional responsibilities;
- roles and differences, if any, between full- and part-time faculty in their participation in institutional policymaking and planning;
- faculty attitudes toward their jobs, their institutions, higher education, and student achievement in general;
- changes in teaching methods, and the impact of new technologies on instructional techniques;
- career and retirement plans;
- differences between those who have instructional responsibilities and those who do not have instructional responsibilities, such as those engaged only in research; and

BEST COPY AVAILABLE

- differences between those with teaching responsibilities but no faculty status and those with teaching responsibilities and faculty status.

3. KEY CONCEPTS

Some key concepts related to NSOPF are described below:

Instructional Faculty/Staff (1998-99).
Faculty—all employees classified by the institution as faculty who were on the institution's payroll as of November 1, 1998. Included as faculty were:

- any individuals who would be reported as "Faculty (Instruction/Research/Public Service)" on the U.S. Department of Education Integrated Postsecondary Education Data System (IPEDS) Fall Staff Survey;
- any individuals with faculty status who would be reported as "Executive, Administrative, and managerial" on the IPEDS Fall Staff Survey, whether or not the person is engaged in any instructional activities; and
- any individuals with faculty status who would be reported as "Other Professionals (Support/Service)" on the IPEDS Fall Staff Survey, whether or not the person is engaged in any instructional activities.

Individuals who would be reported as "Instruction/ Research Assistants" on the IPEDS Fall Staff Survey were excluded.

Instructional Staff—all employees with instructional responsibilities—teaching one or more courses, or advising or supervising students' academic activities (e.g., serving on undergraduate or graduate thesis or dissertation committees, supervising an independent study or one-on-one instructions, etc.)—who may or may not have faculty status. Includes as instructional staff were:

- any individuals with instructional responsibilities during the 1998 Fall Term who would be reported as "Executive, Administrative, and Managerial" on the IPEDS Fall Staff Survey (i.e., A finance officer teaching a class in the business school); and
- any individual with instructional responsibilities during the 1998 Fall Term who would be reported as "Other Professionals (Support/Service)" on the IPEDS Fall Staff Survey.

Individuals who would be reported as "Instruction/ Research Assistants" on the IPEDS Fall Staff Survey were excluded.

Instructional Faculty/Staff (1992–93). All institutional staff (faculty and nonfaculty) whose major regular assignment at the institution (more than 50 percent) was *instruction*. This corresponds to the definition used in the Integrated Postsecondary Education Data System (IPEDS, see chapter 14), which defines faculty (instruction/research) as "all persons whose specific assignments customarily are made for the purpose of conducting instruction, research or public service as a principle activity (or activities) and who hold academic-rank titles of professor, associate professor, assistant professor, instructor, lecturer, or the equivalent of any of these academic ranks. If their principle activity is instructional, [this category also includes] deans, directors, or the equivalent, as well as associate deans, assistant deans and executive officers of academic departments . . ."

A dedicated instructional assignment was not required for an individual to be designated as instructional faculty/staff in the 1992–93 NSOPF. Included in the definition were: (1) administrators whose major responsibility was instruction; (2) individuals with major instructional assignments who had temporary, adjunct, acting, or visiting status; (3) individuals whose major regular assignment was instruction but who had been granted release time for other institutional activities; and (4) individuals whose major regular assignment was instruction but who were on sabbatical leave from the institution. Excluded from this definition were graduate or undergraduate teaching assistants, postdoctoral appointees, temporary replacements for personnel on sabbatical leave, instructional personnel on leave without pay or teaching outside the United States, military personnel who taught only ROTC courses, and instructional personnel supplied by independent contractors.

Noninstructional Faculty (1992–93). All institutional staff who had faculty status but were not counted as instructional faculty since their specific assignment was ***not instruction*** but rather conducting research, performing public service, or carrying out administrative functions of the institution.

Instructional Faculty (1987–88). Those members of the institution's instruction/research staff who were employed full-time or part-time (as defined by the institution) and whose assignment included instruction. Included were: (1) administrators, such as department chairs or deans who held full-time or part-time faculty rank and whose assignment included instruction; (2) regular full-time and part-time instructional faculty; (3) individuals who contributed their instructional services, such as members of religious orders; and (4) instructional faculty on sabbatical leave. Excluded from this definition were teaching assistants; replacements for faculty on sabbatical leave; faculty on leave without pay; and others with adjunct, acting, or visiting appointments.

4. SURVEY DESIGN

Target Population

As of the 1998–99 NSOPF, the target population consists of all public and private, not-for-profit Title IV-participating, 2- and 4-year degree-granting institutions in the 50 states and the District of Columbia that offered programs designed for high school graduates and were open to persons other than employees of the institution, and instructional and noninstructional faculty and staff in these institutions. The 1992–93 and 1987–88 NSOPF institution-level population included postsecondary institutions with accreditation at the college level recognized by the U.S. Department of Education. The 1987–88 NSOPF faculty-level population included only instructional faculty, but the 1987–88 NSOPF also targeted department chairpersons.

Sample Design

The 1998–99 NSOPF used a two-stage sample design, with a sample of 960 institutions in the first stage and a final actual faculty sample of 19,973 faculty.

Institutions were sampled from the 1997–98 Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics (IC) data files and the 1997 and 1995 IPEDS Fall Staffing files. In the institution-level sampling stage, institutions were classified into eight strata by school type, based on their Carnegie Classifications. The eight strata were: (1) public master's (comprehensive) universities and colleges with at least 800 faculty; (2) public master's universities and colleges with fewer than 800 faculty; (3) private master's (comprehensive) universities and colleges; (4) public baccalaureate colleges, including liberal arts colleges, schools of engineering, nursing, and business, teacher's colleges, and other specialized schools; (5) private baccalaureate colleges, including liberal arts colleges, schools of engineering, nursing, and business, teacher's colleges, Bible colleges and theological seminaries, and other specialized schools; (6) medical schools and medical centers; (7) Associates of Arts colleges; and (8) research universities and other doctoral institutions.

In the faculty-level stage of sampling, faculty were grouped into five strata based on their demographic characteristics: (1) Hispanic faculty; (2) Non-Hispanic Black faculty; (3) Asian and Pacific Islander faculty; (4) Full-time female faculty (who were not Hispanic, Black, Asian or Pacific Islander); and (5) All other faculty. Stratifying the faculty in this way allowed for the oversampling of relatively small subpopulations (such as minority group members) to increase the precision of the estimates for these groups. The selection procedure allowed the sample sizes to vary across institutions but minimized the variation in the weights within the staff-level strata: the sampling fractions for each sample institution were made proportional to the institution weight.

To achieve an acceptable response rate for the faculty survey, a subsample of the remaining nonrespondents was drawn for intensive follow up. The design used to carry out this subsampling attempted to reduce the variation in the final cluster sizes by taking a higher fraction of nonrespondents within institutions that had a smaller number of initial faculty selections. Institutions were grouped into three categories: (1) within the sample institutions that had 15 or fewer initial faculty selections; (2) within the institutions with more than 15 initial faculty selections but fewer than 15 respondents at the time of sampling; and (3) within the remaining institutions (all those with at least 15 respondents by the time subsampling was carried out), subsampling was carried out at a lower rate. Altogether the subsample included 3,359 faculty selections. After subsampling, the actual faculty sample size was 19,973.

The 1992–93 NSOPF was conducted with a sample of 974 postsecondary institutions (public and private, not-for-profit 2- and 4-year institutions whose accreditation at the college level was recognized by the U.S. Department of Education) and over 31,000 faculty sampled from institution faculty lists in the second stage. Institutions were selected from IPEDS and then classified into 15 strata by school type, based on their Carnegie Classifications. The strata were: (1) private, other Ph.D. institution (not defined in any other stratum); (2) public, comprehensive; (3) private, comprehensive; (4) public, liberal arts; (5) private, liberal arts; (6) public, medical; (7) private, medical; (8) private, religious; (9) public, 2-year; (10) private, 2-year; (11) public, other type (not defined in any other stratum); (12) private, other type (not defined in any other stratum); (13) public, unknown type; (14) private, unknown type; and (15) public, research; private, research; and public, other Ph.D. institution (not defined in any other stratum). Within each stratum, the institutions were further sorted by school size. Of the 962 eligible institutions, 817 institutions (85 percent) provided lists of faculty. The selection of faculty within each institution was random except for the oversampling of the following groups: Blacks (both non-Hispanics and Hispanics); Asians/Pacific Islanders; faculty in disciplines specified by the National Endowment for the Humanities; and full-time female faculty.

The 1987–88 NSOPF was conducted with a sample of 480 institutions (including 2-year, 4-year, doctoral-granting, and other colleges and universities), over 11,000 faculty, and more than 3,000 department chairpersons. Institutions were sampled from the 1987 IPEDS universe and were stratified by modified Carnegie Classifications and size (faculty counts). These strata were (1) public, research; (2) private, research; (3) public, other Ph.D. institution (not defined in any other stratum); (4) private, other Ph.D. institution (not defined in any other stratum); (5) public, comprehensive; (6) private, comprehensive; (7) liberal arts; (8) public, 2-year; (9) private, 2-year; (10) religious; (11) medical; and (12) "other" schools (not defined in any other stratum). Within each stratum, institutions were randomly selected. Of the 480 institutions selected, 449 (94 percent) agreed to participate and provided lists of their faculty and department chairpersons. Within 4-year institutions, faculty and department chairpersons were stratified by program area and randomly sampled within each stratum; within 2-year institutions, simple random samples of faculty and department chairpersons were selected; and within specialized institutions (religious, medical, etc.), faculty samples were randomly selected (department chairpersons were not sampled). At all institutions, faculty were also stratified on the basis of employment status—full-time and part-time. Note that teaching assistants and teaching fellows were excluded in the 1987–88 NSOPF.

Data Collection and Processing

The 1998–99 NSOPF allowed sample members to complete a paper self-administered questionnaire and mail it back or to complete the questionnaire via the Internet. Follow-up activities included e-mails, telephone prompting, and, for nonresponding faculty, computer-assisted telephone interviewing (CATI). As part of the study, an experiment was conducted to determine if small financial incentives could increase use of the web-based version of the questionnaire. Previously, NSOPF was a mailout/mailback survey with telephone follow up. The 1987–88 NSOPF was conducted by SRI International, the 1992–93 NSOPF by the National Opinion Research Center (NORC) at the University of Chicago, and the 1998–99 NSOPF by The Gallup Organization.

Reference dates. Most of the information collected in the NSOPF pertains to the Fall Term of the academic year surveyed. For the 1998–99 NSOPF, the Fall Term was defined as the academic term containing November 1, 1998. The Institution Survey also asked about the number of full-time faculty/staff hired since the 1991 Fall Term; the number of tenured and tenure-track faculty in both the 1997 and 1998 Fall Terms; the consideration and granting of tenure during the 1997–98 academic year; and the number of faculty, granting of tenure and early/phased retirement in the previous 5 years. The 1998–99 NSOPF Faculty Survey asked faculty members about their gross compensation, household income, number in household, and number of dependents in calendar year 1998; their presentations and publications in the last 2 years; and the likelihood of leaving their current job in the next 3 years (and the reasons). Similarly, the 1992–93 and the 1987–88 NSOPF requested most information for the 1992 and 1987 Fall Term, respectively, but included some questions requiring retrospective or prospective responses.

Data collection. The 1998–99 NSOPF institution and faculty data collection offered both a paper and a web version of the questionnaire, with telephone (including computer-assisted telephone interviews) and e-mail follow up. The data collection procedure started with a prenotification letter to the institution's CAO to introduce the CAO to the study, and secure the name of an appropriate individual to serve as Institution Coordinator (i.e., the individual at the school who would be responsible for the completing the data request). The data collection packet was then mailed directly to the Coordinator. The packet contained both the Institution Questionnaire and the list collection packet. The Coordinator was asked to complete and return all materials at the same time. The mailing was timed to immediately precede the November 1, 1998, reference date for the fall term.

The field period for the 1998–99 NSOPF Faculty Survey extended from February 1999 through March 2000. Questionnaires were mailed to faculty in batches or waves, as lists of faculty and instructional staff were received, processed, and sampled. Questionnaires were accompanied by a letter that provided the web address and a personal identification (PIN) code to be used to access the web questionnaire. The first wave of questionnaires was mailed on February 4, 1999; the seventh and final wave was mailed on December 1, 1999. Faculty sample members in each wave received a coordinated series of mail, e-mail, and telephone follow up. Mail follow up for nonrespondents included a postcard and up to four questionnaire re-mailings; these were mailed to the home address of the faculty member if provided by the institution. E-mail prompts were sent to all faculty for whom an e-mail address was provided. Faculty received as many as six e-mail prompts. Telephone follow up consisted of initial prompts to complete the mail or web questionnaire. A CATI was scheduled for nonrespondents to the mail, e-mail, and telephone prompts.

The following efforts were made for the 1992–93 NSOPF Institution Survey: initial questionnaire mailing, postcard prompting, second questionnaire mailing, second postcard prompting, telephone prompting, third questionnaire mailing, and telephone interviewing. Similarly, the data for the 1992–93 NSOPF Faculty Survey were collected through an initial questionnaire mailing, postcard prompting, second questionnaire mailing, third questionnaire mailing, telephone prompting, and CATI. For both surveys, institutions and faculty who missed critical items and/or had inconsistent or out-of-range responses were identified for data retrieval. Extra telephone calls were made to retrieve these data. Data collection procedures for the 1987–88 NSOPF involved three mailouts for both the Institution Survey and the Department Chairperson Survey, and two mailouts and one CATI interview for the Faculty Survey.

Data processing. The three modes of questionnaire administration in the 1998–99 NSOPF each required separate systems for data capture. All self-administered paper questionnaires were optically scanned. The system was programmed so that each character was read and assigned a confidence level. All characters with less than a 100 percent confidence level were automatically sent to an operator for manual verification. The contractor verified the work of each operator and the recognition engines on each batch of every questionnaire to ensure that the quality assurance system was working properly. Also, 100 percent of written out responses (as opposed to check marks) were manually verified.

Each web respondent was assigned a unique access code, and respondents without a valid access code were not permitted to enter the web site. A respondent could return to the survey web site at a later time to complete a survey that was left unfinished in an earlier session. When respondents entered the web site using the access code, they were immediately taken to the same point in the survey item sequence that they had reached during their previous session. If a respondent, re-using an access code, returned to the web site at a later time after completing the survey in a previous session, they were not allowed access to the completed web survey data record. Responses to all web-administered questionnaires underwent data editing, imputation, and analysis.

All telephone interviews used CATI technology. The CATI program was altered from the paper questionnaire to ensure valid codes, perform skip patterns automatically, and make inter-item consistency checks where appropriate. The quality control program for CATI interviewing included project specific training of interviewers, regular evaluation of interviewers by interviewing supervisors, and regular monitoring of interviewers.

In the 1992–93 NSOPF, both computer-assisted data entry (CADE) and CATI were used. The CADE/CATI systems were designed to ensure that all entries conformed to valid ranges of codes; enforced skip patterns automatically; conducted inter-item consistency checks where appropriate; and displayed the full question and answer texts for verbatim responses. As part of the statistical quality control program, 100 percent verification was conducted on a randomly selected subsample of 10 percent of all institution and faculty questionnaires entered in CADE. The error rate was less than 0.5 percent for all items keyed. Quality assurance for CATI faculty interviews consisted of random online monitoring by supervisors.

Coding of institution questionnaires. The 1998–99 NSOPF Institution Questionnaire had few "other specify" questions, and no coding was performed. For the 1992–93 NSOPF, coding was performed for verbatim definitions of full-time and part-time faculty (both instructional and noninstructional) and for permanent and temporary faculty. Six other institution questionnaire items were eligible for verbatim or "other specify" responses. Only two provided consistent verbatim responses; these questions asked for a description of "any other actions" taken to lower the percentage of tenured faculty for full-time instructional and for full-time noninstructional faculty.

Coding of faculty questionnaires. Four categories of open-ended questions required coding in the 1998–99 Faculty Questionnaire: academic discipline, IPEDS codes, country of educational institution or birth, and "other specify" questions. Academic discipline was partially precoded by either the respondent or the interviewer. All other coding was done as a post-processing step. Many open-ended responses were coded automatically using SAS software, but county codes, "other specify," and verbatim text were hand-coded by project staff.

For the 1992–93 NSOPF, coding was conducted using a computer-assisted coding system. Coding of academic discipline was performed online during interviewing or data entry. All other faculty questionnaire coding was performed after other processing. Coding was performed for the following: academic discipline for the respondent's principal teaching field, principal area of research, degree fields, and courses taught (using codes supplied with the survey); institutions that awarded academic degrees (using IPEDS codes); country of birth and/or citizenship; country of foreign institution for institutions that could not be coded within the IPEDS codeframe (using codes compiled for the 1987–88 NSOPF); and "other specify" and verbatim text (in most cases, coded to existing codes).

Editing. Besides the procedures described above under "Processing," the following editing procedures were implemented for the 1998–99 NSOPF:

- *Menu items.* Several procedures were instituted to clean responses to questions that had sub-items listed where the respondent was asked to give a response for each sub-item. If the main question had an "NA" (Not Applicable) check box and that box was checked, all of the sub-items were set to a value of "no" or "zero" depending on the wording of the question. If the respondent had filled out one or more of the sub-items with a "yes" response or a positive number but had left other sub-items blank, the missing sub-items were set to "no," "zero," or "don't know" depending on the question wording. If all sub-items were missing and there was no "NA" box, or the "NA" box was not checked, the case was flagged and the data values were imputed for that question.
- *Inter-item consistency checks.* Many types of inter-item consistency checks were performed on the data. One procedure was to check groups of related items for internal consistency and to make adjustments to make them consistent. Another procedure checked "NA" boxes. If the respondent had checked the "NA" box for a question but had filled in any of the sub-items for that question the "NA" box was set to blank. A third procedure was to check filter items for which more detail was sought in a follow-up open-ended or closed-ended question. If detail was provided, then the filter question was checked to make sure the appropriate response was recorded.
- *Percent items.* All items where respondents were asked to give a percentage were checked to make sure they summed to 100 percent. The editing program also looked for any numbers between 0 and 1 to make sure that respondents did not fill in the question with a decimal rather than a percentage. All fractions of a percent were rounded to the nearest whole percent.

Estimation Methods

Weighting was used in NSOPF to adjust for sampling and unit nonresponse at both the institution and faculty levels. Imputation was performed to compensate for item nonresponse.

Weighting. Three weights were computed for the 1998–99 NSOPF: full-sample institution weights, full-sample faculty weights, and a contextual weight (to be used in "contextual" analyses that simultaneously include variables drawn from the faculty and institution questionnaires). The formulas representing the construction of each of these weights are provided in the *1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report* (NCES 2001–151).

The weighting of the 1992–93 and 1987–88 NSOPFs is described below.

1992–93 NSOPF. Three weights were computed for the 1992–93 NSOPF sample—first-stage institution weights, final institution weights, and final faculty weights. The first-stage institution weights accounted for the institutions that participated in the study by submitting a faculty sampling list that allowed faculty members to be sampled. The two final weights—weights for the sample faculty and institution weights for those institutions that returned Institution Surveys—were adjusted for nonresponse. The final faculty weights were poststratified to the "best" estimates of the number of faculty. The "best" estimates were derived following reconciliation and verification through recontact with a subset of institutions that had discrepancies of 10 percent or greater between the total number enumerated on the faculty list used for sampling and the total number reported on the Institution Survey. For more information on the reconciliation effort, refer to "Measurement error" in section 5 of this chapter. For more information on the calculation of the "best" estimates of faculty, refer to the *1993 National Study of Postsecondary Faculty: Methodology Report* (NCES 97–467).

1987–88 NSOPF. The 1987–88 NSOPF sample was weighted to produce national estimates of institutions, faculty, and department chairpersons by using weights designed to adjust for differential probabilities of selection and nonresponse at the institution, faculty, and department chairperson levels. The sample weights for institutions were calculated as the inverse of the probability of selection, based on the number of institutions in each size substratum. Sample weights were adjusted to account for nonresponse by multiplying the sample weights by the reciprocal of the response rate. Sample weights for the 1987–88 faculty summed to the total number of faculty in the IPEDS universe of institutions, as projected from the lists of total faculty provided by participating institutions. Sample weights accounted for two levels of nonresponse, one for nonparticipating institutions and the other for nonresponding faculty. Sample weights for the departments in the 1987–88 NSOPF summed to the estimated total number of departments in the IPEDS universe of institutions. Sample weights accounted for nonresponse of nonparticipating institutions and nonresponding department chairpersons.

Imputation. Data imputation for the 1998–99 NSOPF Faculty Questionnaire was performed in four steps.

(1) *Logical imputation.* The logical imputation was conducted during the data cleaning steps as explained under "Processing."

(2) *Cold deck.* Missing responses were filled in with data from the sample frame whenever the relevant data were available.

(3) *Sequential hot deck.* Nonmissing values were selected from "sequential nearest neighbors" within the imputation class. All questions that were categorical and had more than 16 categories were imputed with this method.

(4) *Regression type.* This procedure employed SAS PROC IMPUTE. All items that were still missing after the logical, cold-deck, and hot-deck imputation procedures were imputed with this method. Project staff selected the independent variables by first looking through the questionnaire for logically related items and then by conducting a correlation analysis of the questions against each other to find the top correlates for each item.

Data imputation for the Institution Questionnaire used three methods. Logical imputation was also performed in the cleaning steps described under "Processing."

(1) *Within-class mean.* The missing value was replaced with the mean of all nonmissing cases within the imputation class. Continuous variables with less than 5 percent missing were imputed with this method.

(2) *Within-class random frequency.* The missing value was replaced by a random draw from the possible responses based on the observed frequency of nonmissing responses within the imputation class. All categorical questions were imputed with this method, since all categorical items had less than 5 percent missing data.

(3) *Hot deck.* As with the faculty imputation, this method selected nonmissing values from the "sequential nearest neighbor" within the imputation class. Any questions that were continuous variables and had more than 5 percent missing cases were imputed with this method.

For a small number of items, special procedures were used. See the *1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report* (NCES 2001–151).

In the 1992–93 NSOPF, two imputation methods were used for the Faculty Survey—PROC IMPUTE and the "sequential nearest neighbor" hot-deck method. PROC IMPUTE alone was used for the Institution Survey. All imputation was followed by a final series of cleaning passes that resulted in generally clean and logically consistent data. Some residual inconsistencies between different data elements remained in situations where it was impossible to resolve the ambiguity as reported by the respondent.

Although the 1987–88 NSOPF consisted of three surveys, imputations were only performed for faculty item nonresponse. The ***within-cell random imputation method*** was used to fill in most Faculty Survey items that had missing data.

Recent Changes

Data from the 1998–99 NSOPF administration will be released in 2001. As in 1992–93, the 1998–99 NSOPF was limited to surveys of institutions and faculty/instructional staff. It allows comparisons to be made over time and also examines critical issues surrounding faculty and instructional staff that have developed since the first two studies. While some aspects remained the same as in the 1992–93 NSOPF, others changed. These include providing a booklet of instructions to the Institutional Coordinator at each institution, separating mailings sent to the CAOs and Institutional Coordinators, requesting faculty lists and Institution Surveys at the same time, personalizing mailings, providing a glossary of terms with the surveys, providing consistent instructions, changing the reference date for faculty employment to November 1, making surveys available on the Internet, utilizing e-mail prompts to institutions and faculty, providing an NSOPF 1998–99 e-mail address for respondents, optically scanning survey responses, and offering institutions a peer report of findings.

Future Plans

NSOPF will be conducted again in the 2003–04 academic year.

5. DATA QUALITY AND COMPARABILITY

The 1998–99 NSOPF included procedures for both minimizing and measuring nonsampling errors. A field test was performed before the 1998–99 NSOPF, and quality control activities continued during interviewer training, data collection, and processing of survey data.

Sampling Error

Standard errors for all NSOPF data can be computed using a technique known as Taylor Series approximation. Individuals opting to calculate variances with the Taylor Series approximation method should use a "with replacement" type of variance formula. Specialized computer programs, such as SUDAAN, calculate variances with the Taylor Series approximation method. The Data Analysis System (DAS) available on CD-ROM calculates variances using the Taylor Series method.

Replicate weights are provided on the NSOPF data files (64 sets of replicates in the 1998–99 NSOPF and 32 replicate weights in the 1992–93 NSOPF). These weights implement the balanced half-sample (BHS) method of variance estimation. They have been created to handle the certainty stratum and to incorporate finite population correction factors for each of the 14 noncertainty strata. Two widely available software packages, WesVar and PC CARP, have capabilities to use replicate weights to estimate variances.

Analysts should be cautious about use of BHS-estimated variances that relate to one stratum or to a group of two or three strata. Such variance estimates may be based upon far fewer than the number of replicates; thus, the variance of the variance estimator may be large. Analysts who use either the restricted-use faculty file or the institution file should also be cautious about cross-classifying data so deeply that the resulting estimates are based upon a very small number of observations. Analysts should interpret the accuracy of the NSOPF statistics in light of estimated standard errors and the small sample sizes.

Nonsampling Error

To minimize the potential for nonsampling errors, the 1998–99 NSOPF Institution and Faculty Surveys (as well as the sample design, data collection, and data processing procedures) were field-tested with a national probability sample of 162 postsecondary institutions and 512 faculty members. Four methodological experiments were conducted as part of the field test. These included experiments to increase unit response rates, speed the return of mail questionnaires, increase data quality, and improve the overall efficiency of the data collection process. The experiments involved the use of prenotification, prioritized mail, a streamlined instrument, and the timing of CATI attempts. Another focus of the field test was the effort to reduce discrepancies between the faculty counts derived from the list of faculty provided by each institution and those provided in the Institution Questionnaire. Changes introduced to reduce discrepancies included providing clearer definitions of faculty eligibility (with consistency across forms and questionnaires) and collecting list and institution questionnaire data simultaneously with the objective of increasing the probability that both forms would be completed by the same individual and evidence fewer inconsistencies.

During the 1992–93 NSOPF field test, a subsample of faculty respondents were reinterviewed to evaluate reliability. In addition, an extensive item nonresponse analysis of the field-tested surveys was conducted, followed by additional evaluation of the instruments and survey procedures. An item nonresponse analysis was also conducted for the full-scale surveys. Later, in 1996, NCES analyzed discrepancies in the 1992–93 faculty counts, conducting a retrieval, verification, and reconciliation effort to resolve problems.

Coverage error. Because the IPEDS universe is the institutional frame for the NSOPF, coverage of institutions is complete. However, there are concerns about the coverage of faculty and instructional staff. In an effort to decrease the discrepancies in faculty counts noticed in the 1992–93 NSOPF, the 1998–99 NSOPF asked the Institution Coordinators to provide counts of full- and part-time faculty and instructional staff at their institutions as of November 1, 1998, the same reference period used for the IPEDS Fall Staff Survey, asked them to return both the faculty list and the Institution Questionnaire at the same time, and—giving them explicit warnings about potential undercounts of faculty—asked them to ensure that the counts provided in the list and questionnaire were consistent. These efforts appear to have worked, since 73 percent of institutions provided questionnaire and list data that exhibited discrepancies of less than 10 percent, an improvement of 31 percentage points since 1993.

In the 1992–93 NSOPF Institution Survey, a discrepancy between the faculty counts and those provided on faculty lists by institutions at the beginning of the sampling process necessitated the "best estimates" correction to the 1992–93 NSOPF faculty population estimates, as described earlier in section 4, Weighting.

Nonresponse error.
Unit nonresponse. Unit response rates have been similar over NSOPF administrations. (See table below.) Note that the overall faculty response rates are the percentage of faculty responding in institutions that provided faculty lists for sampling.

Table 5. Summary of weighted response rates for selected NSOPF surveys

Questionnaire	List participation rate	Questionnaire response rate	Overall
NSOPF 1992–93			
Institution	†	93.6	93.6
Faculty	84.4	83.4	70.4
NSOPF 1998–99			
Institution	†	92.8	92.8
Faculty	88.4	83.0	73.4

†Not applicable
SOURCE: Abraham, Steiger, Montgomery, Kuhr, Tourangeau, Montgomery, and Chattopadhyay, *1999 National Study of Postsecondary Faculty (NSOPF:99)* (NCES 2001–151). Selfa, Suter, Myers, Koch, Johnson, Zahs, Kuhr, and Abraham, *1993 National Study of Postsecondary Faculty (NSOPF) Methodology Report* (NCES 97–467).

In the 1987–88 NSOPF, the unweighted response rates (weighted response rates are not available) were: 88.3 percent for the Institution Survey; 76.1 percent for the Faculty Survey, and 80.1 percent for the Department Chairperson Survey.

Item nonresponse. For the 1998–99 NSOPF Institution Questionnaire, the mean item nonresponse rate was 4.3 percent (unweighted). Twenty-one items had item nonresponse rates greater than 10 percent; one item had a nonresponse rate greater than 20 percent. The situation is complicated for the Faculty Questionnaire because an abbreviated questionnaire (containing 202 of the total 369 items in the full questionnaire) was administered to most CATI respondents. For all questions the average nonresponse was 19.2 percent; with just the 202 items on the abbreviated questionnaire, the average nonresponse was 15.5 percent. For further details on item nonresponse, see the *1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report* (NCES 2001–151).

For the 1992–93 Institution Survey, the mean item nonresponse rate was 10.1 percent, with the level of nonresponse increasing in the latter parts of the survey. For the Faculty Survey, the mean item nonresponse rate was 10.3 percent.

Measurement error. For the 1998–99 NSOPF, NCES conducted an intensive follow up with 234 (28.6 percent of participating) institutions whose reports exhibited a variance of 5 percent or more between the list and questionnaire counts overall, or between the two part-time counts. The NSOPF survey system has experienced discrepancies in faculty counts among IPEDS, institution questionnaire, and the list of faculty across all cycles of the study. Even though the identical information is requested on the questionnaire as on the list (i.e., a count of all full-time and part-time faculty and instructional staff as of November 1, 1998), institutions have continued to provide discrepant faculty data to NSOPF requests. As in 1993, large discrepancies tend to be concentrated among smaller institutions, and 2-year institutions. Undercounting of part-time faculty and instructional staff without faculty status on the list remains the primary reason for the majority of these discrepancies.

However, procedures implemented in NSOPF:99 improved the consistency of the list and questionnaire counts when compared to previous cycles of NSOPF. The percent of institutions providing list and questionnaire data that had less than a 10 percent discrepancy increased from 42 percent in NSOPF-93 to 73 percent in NSOPF:99. A total of 43 percent provided identical data on the list and questionnaire in NSOPF:99 (compared to only 2.4 percent in 1993). Moreover, schools providing identical list and questionnaire data were shown to have provided more accurate and complete data on both the lists and questionnaire. These findings suggest that the changed procedures that were introduced in the 1998 field test and NSOPF:99 resulted in more accurate counts of faculty and instructional staff. Institutions may also be in a better position to respond to these requests for data. Their accumulated experience in handling NSOPF and IPEDS (and other survey) requests, their adoption of better reporting systems, more flexible computing systems and staff, and a general willingness to provide the information are probably also a factor in their ability to provide more consistent faculty counts although data to support these assertions are not available. For more detail, see *1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report* (NCES 2001–151).

NCES conducted three studies to examine possible measurement errors in the 1992–93 NSOPF: (1) a reinterview study of selected faculty questionnaire items, conducted after the field test; (2) a discrepancy and trends analysis of faculty counts in the full survey; and (3) a retrieval, verification, and reconciliation effort involving recontact of institutions. For detail on these studies, see *Measurement Error Studies at the National Center for Education Statistics* (NCES 97–464) and *1993 National Study of Postsecondary Faculty: Methodology Report* (NCES 97–467).

Reinterview study. A reliabilility reinterview study was conducted after the 1992–93 NSOPF field test for the purpose of identifying faculty questionnaire items that yielded low quality data and the item characteristics that caused problems, thus providing a basis for revising the questionnaire items prior to implementation of the full-scale survey. The analysis of the reinterview items was presented by item type—categorical or continuous variables—rather than by subject area. The level of consistency between the field test responses and the reinterview responses was relatively high: a 70 percent consistency for most of the categorical questions and a 0.7 correlation for most of the continuous variables. A detailed analysis of the question on employment sector of last main job was conducted because it showed the highest percentage of inconsistent responses (28 percent) and the highest inconsistency index (36.0). It was concluded that the large number of response categories and the involvement of some faculty in more than one job sector were plausible reasons for the high inconsistency rate. The items with the lowest correlations were those asking for retrospective reporting of numbers that were small fractions of dollars or hours and those asking for summary statistics on activities that were likely to fluctuate over time—the types of questions shown to be unreliable in past studies.

Discrepancy and trends analysis of faculty counts. This analysis compared discrepancies between different types of institutions to identify systematic sources of discrepancies in faculty estimates between the faculty list counts provided by the institution for sampling and faculty counts reported in the Institution Questionnaire. The investigation found that list estimates tended to exceed questionnaire estimates in large institutions, in institutions with medical components, and in private schools. Questionnaire estimates tended to be higher in smaller institutions, in institutions without medical components, and in public schools. Institutions supplied much higher questionnaire estimates for part-time faculty than list estimates. Faculty lists submitted early in the list collection process showed little difference in the magnitude of questionnaire/list discrepancies from faculty lists submitted later in the process.

Retrieval, verification, and reconciliation. This effort involved recontacting 509 institutions: 450 institutions (more than half of all institutions) whose questionnaire estimate of total faculty differed from the institution's list estimate by 10 percent or more, and an additional 59 institutions NCES designated as operating medical schools or hospitals. All institutions employing health sciences faculty and participating in the 1992–93 NSOPF were selected for recontact.

NCES accepted the reconciled estimates obtained in this study as the true numbers of faculty. More than one-half (56.9 percent) of the recontacted institutions identified the questionnaire teacher estimate as the most accurate response, while 24.8 percent identified the list estimate as the most accurate. Another 11.4 percent of the institutions provided a new estimate; 1 percent indicated that their IPEDS teacher estimate was the most accurate estimate; and 5.9 percent could not verify any of the estimates and thus accepted the original list estimate.

The majority of discrepancies in faculty counts resulted from the exclusion of some full- or part-time faculty from the list or questionnaire. Another factor was the time interval between the date the list was compiled and the date the questionnaire was completed. Downsizing also affected faculty counts at several institutions. Some of the reasons for the discrepancies were unexpected. For example, some institutions provided "full-time equivalents" (FTEs) on the Institution Survey instead of an actual headcount of part-time faculty.

Sometimes part-time faculty were overreported—often a result of confusion over the pool of part-time and temporary staff employed by or available to the institution during the course of the academic year versus the number actually employed during the fall semester. Another reason given for overreporting of part-time faculty was an inability to distinguish honorary/unpaid part-time faculty from paid faculty and teaching staff. This study also confirmed that a small number of institutions excluded medical school faculty from their lists of faculty. In those cases, the institutions considered their medical schools separate from their main campuses.

While these results indicate that there may have been some bias in the 1992–93 NSOPF sample, no measure of the potential bias, such as the net difference rate, was computed. Instead, the reconciliation prompted NCES to apply a poststratification adjustment to the estimates based entirely on the "best" estimates obtained during the reinterview study described above. Problems with health science estimates, however, could only be partly rectified by the creation of new "best" estimates. For more information on the calculation of the "best" estimates and further discussion of the health science estimates, refer to the *1993 National Study of Postsecondary Faculty: Methodology Report* (NCES 97–467).

Data Comparability

The comparison of 1998–99 NSOPF faculty questionnaire data with 1992-93 NSOPF "best estimates" shows, overall, continuing growth in both full- and part-time faculty. Faculty growth varies widely by strata, however, and some strata report fewer faculty than in 1993 (e.g., public comprehensive faculty, private medical faculty) while others remain virtually unchanged (e.g., public and private 2-year faculty). In some instances, changes in individual strata may simply reflect changes in the institutional composition of individual strata since 1993, as well as shifts in the numbers of faculty employed at institutions within each stratum. (Moreover, some institutions included in the 1993 sample may have changed classification.) Despite shifts in the faculty counts of individual strata, the percentages of full and part-time faculty in each strata are closely comparable to what was reported as a "best estimate" in 1993.

Design changes. Each succeeding cycle of NSOPF has expanded the information base about faculty. The 1998–99 NSOPF is designed both to facilitate comparisons over time and to examine new faculty-related issues that have emerged since the 1992–93 study. The 1998–99 sample was designed to allow detailed comparisons and high levels of precision at both the institution and faculty levels. In the 1998–99 study, the definition of institutions changed to match the IPEDS definition. Since the 1992–93 study, the operant definition of "faculty" for NSOPF has included instructional faculty, noninstructional faculty and instructional personnel without faculty status.

The 1998–99 and 1992–93 NSOPF consisted of two surveys: an Institution Survey and a Faculty Survey. The 1987–88 NSOPF included a Department Chairperson Survey in addition to the Institution Survey and the Faculty Survey.

Definitional differences. *Comparisons among the three cycles must be made cautiously because the respondents in each cycle were different.* On the institution level, the 1998–99 NSOPF sample consists of all public and private, not-for-profit Title IV-participating, degree-granting institutions in the 50 states and the District of Columbia. This change was made so that the NSOPF sampling universe conformed with that of IPEDS. In previous rounds of the study, the sample consisted of public and private not-for-profit 2- and 4-year (and above) higher education institutions.

The definition of faculty and instructional staff for each NSOPF cycle is given under key concepts. On the design level, note that the 1998–99 and 1992–93 NSOPF requested a listing of *all faculty (instructional and noninstructional) and instructional staff* from the institutions for purposes of sampling. For the 1987–88 NSOPF, institutions were asked to provide only the names of *instructional faculty*. Although not specifically stated, NCES expected that institutions would provide information on instructional staff as well. The term faculty was used generically. There is no way of knowing how many institutions that had instructional staff as well as instructional faculty provided names for both. Each institution was allowed to make its own decision about which faculty members belonged in the sample, thereby creating a situation that does not allow researchers to precisely match the *de facto* sample definition used by institutions in the 1987–88 NSOPF.

Content changes. For the purpose of trend analysis, as many of the 1992–93 items as were relevant and feasible were retained in the 1998–99 questionnaires. However, this goal had to be balanced with the need to address recent policy issues. In the Institution Questionnaire, 17 items were revised from the 1992–93 questionnaire, and 7 new items were added. In the Faculty Questionnaire, 44 items were revised, and 32 new items were added.

Comparisons with other surveys. Comparisons of 1992–93 NSOPF salary estimates with salary estimates from IPEDS and from the American Association of University Professors indicate that NSOPF data are consistent with these other sources. Most differences are relatively small and can be easily explained by methodological differences between the studies. The NSOPF estimates are based on self-reports of individuals, whereas the other two studies rely on institutional reports of salary means for the entire institution.

However, the reader should be aware of differences in faculty definitions between NSOPF and IPEDS. The differences between the IPEDS definition and NSOPF's is that a person in IPEDS has to be categorized according to their primary responsibility (administrator, faculty, or other professional); whereas, in NSOPF it is possible to categorize according to any of their responsibilities.

Because NSOPF includes all faculty and instructional staff, it is possible for an "other professional" to have instructional responsibilities and/or be a faculty member, and it is also possible for an administrator to have instructional responsibilities and/or be a faculty member. Therefore, NSOPF includes all faculty under IPEDS, some of the administrators under IPEDS, and some of the other professionals under IPEDS.

6. CONTACT INFORMATION

For content information on the NSOPF, contact:

Aurora M. D'Amico
Phone: (202) 502–7334
E-mail: aurora.d'amico@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

1988 National Survey of Postsecondary Faculty Methodology Report, by S.H. Russell, M.P. Hancock, and C. Williamson. Menlo Park, CA: 1990.

1999 National Study of Postsecondary Faculty (NSOPF:99) Methodology Report, NCES 2001–151, by S.Y. Abraham, D.M. Steiger, M. Montgomery, B.D. Kuhr, R. Tourangeau, B. Montgomery, and M. Chattopadhyay. Washington, DC: 2001.

1993 National Study of Postsecondary Faculty (NSOPF-93): Methodology Report, NCES 97–467, by L.A. Selfa, N. Suter, S. Myers, S. Koch, R.A. Johnson, D.A. Zahs, B.D. Kuhr, and S.Y. Abraham. Washington, DC: 1997.

1992–93 National Study of Postsecondary Faculty Field Test Report, NCES 93–390, by S.Y. Abraham, N.A. Suter, B.D. Spencer, R.A. Johnson, D.A. Zahs, and S.L. Myers. Washington, DC: 1994.

Survey Design

Guidebook for Imputation of Missing Data, by American Institutes for Research. Washington, DC: 1980.

1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report, NCES Working Paper 2000–01, by S.Y. Abraham, D.M. Steiger, R. Tourangeau, B.D. Kuhr, B. Wells, and Y. Yang. Washington, DC: 2000.

Data Quality and Comparability

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, M. Saba, and S. Kaufman. Washington, DC: 1997.

Strategies for Improving Accuracy of Postsecondary Faculty Lists, NCES Working Paper 97–26, by S. Salvucci, R. Bureika, G. Carter, D. Ghosh, M. Reiser, and S. Wenck. Washington, DC: 1997.

Chapter 16: National Postsecondary Student Aid Study (NPSAS)

1. OVERVIEW

The National Postsecondary Student Aid Study (NPSAS) is a comprehensive nationwide study conducted by NCES to determine how students and their families pay for postsecondary education. It is designed to address policy questions resulting from the rapid growth of financial aid programs and the succession of changes in financial aid program policies since 1986. The first NPSAS was conducted during the 1986–87 academic year. The fifth in the series was administered during the 1999–2000 academic year.

NPSAS is based on a nationally representative sample of all students in postsecondary education institutions in the 50 states, the District of Columbia, and Puerto Rico. Institutions may be public or private, and they may be less than 2-year schools, community colleges (2–3 years), 4-year colleges, or major universities with graduate-level programs. Study participants include students who receive financial aid as well as those who do not. NPSAS data are obtained from administrative records of student financial aid, interviews with students, and interviews with a subsample of parents. Information has been gathered on more than 55,000 students in each study cycle.

NPSAS also provides baseline data for two longitudinal studies: the Beginning Postsecondary Students (BPS) Longitudinal Study and the Baccalaureate and Beyond (B&B) Longitudinal Study. (See chapters 17 and 18.) The 1990 and 1996 NPSAS studies served as baselines for BPS cohorts; the 1993 and 2000 NPSAS studies were the baseline for the two B&B cohorts.

SAMPLE SURVEY OF POSTSECONDARY INSTITUTIONS AND STUDENTS; CONDUCTED EVERY 3–4 YEARS

NPSAS collects information from:

- Student institutional record abstracts
- Department of Education administrative records
- Student interviews
- Parent interviews

Purpose

To produce reliable national estimates of characteristics related to financial aid for postsecondary students. The study also describes demographic and other characteristics of those enrolled. The study focuses on three topics: (1) how students and their families finance postsecondary education; (2) how the process of financial aid works, in terms of both who applies and who receives aid; and (3) the effects of financial aid on students and their families.

Components

There are four components to NPSAS, described below.

Student Record Abstract. The following information on students is obtained from institutional records: year in school; major field of study; type and control of institution; attendance status; tuition and fees; admission test scores; financial aid awards; cost of attendance; student budget information and expected family contribution for aided students; grade point average; age; and date first enrolled. An appointed Institutional Coordinator or a field data collector extracts the information from student records and enters it into a customized computer-assisted data entry system.

Department of Education Administrative Records. Beginning in 1995–96, the following information has been collected from Department of Education administrative records on financial aid applications and loans: types and amounts of federal financial aid received; cumulative loan amounts from the National Student Loan Data System; and loan repayment status.

Student Interview. Telephone interviews with students provide data on level (undergraduate, graduate, first-professional); major field of study; financial aid at other schools attended during the year; other sources of financial support; reasons for selecting the school they are attending; current marital status; age; race/ethnicity; sex; highest degree expected; employment and income; voting in recent elections; and community service.

Parent Interview. Telephone interviews with a limited sample of students' parents (through 1995–96) collect supplemental data, including parents' marital status; age; highest level of education achieved; income; amount of financial support provided to children; types of financing used to pay child's educational expenses; and occupation and industry. No parent interviews are planned after 1995–96.

Periodicity

Triennial from 1986–87 through 1995–96, and quadrennial beginning in 1999–2000.

2. USES OF DATA

The goal of the NPSAS study is to identify institutional, student, and family characteristics related to participation in financial aid programs. Federal policymakers use NPSAS data to determine future federal policy concerning student financial aid. With these data, it is possible to analyze special population enrollments in postsecondary education, including students with disabilities, racial and ethnic minorities, students taking remedial/developmental courses, students from families with low incomes, and older students. The distribution of students by major field of study can also be examined. Fields of particular interest are mathematics, science, and engineering, as well as teacher preparation and health studies. Data can also be generated on factors associated with choice of postsecondary institution, participation in postsecondary vocational education, parental support for postsecondary education, and occupational and educational aspirations.

It is important that statistical analyses be conducted using software that properly accounts for the complex sampling design of NPSAS. NCES has developed a software tool called the Data Analysis System (DAS) for analysis of complex survey data. For information on other software packages and statistical strategies useful for analysis of complex survey data, see appendix F of *National Postsecondary Student Aid Study 1995–96 (NPSAS:96), Methodology Report* (NCES 98–073).

3. KEY CONCEPTS

Described below are several key concepts relevant to financial assistance for postsecondary education. For additional NPSAS terms, refer to the glossaries in published statistical analysis reports and database documentation.

Institution Type. A derived variable that combines information on the level and control of the NPSAS institution. Institution level concerns the institution's length of program and highest degree offering and is defined as less than 2-year, 2- to 3-year, 4-year nondoctorate, or 4-year doctorate (including first-professional degree). Institution control concerns the source of revenue and control of operations and is defined as public, private not-for-profit, or private for-profit.

Attendance Pattern. A student's intensity and persistence of attendance during the NPSAS year. Intensity refers to the student's full- or part-time attendance while enrolled. Persistence refers to the number of months a student is enrolled during the year. Students are considered to be enrolled for a full year if they are enrolled 8 or more months during the year. Months do not have to be contiguous or at the same institution, and students do not have to be enrolled for a full month to be considered enrolled for that month. In surveys prior to the 1995–96 NPSAS, full year was defined as 9 or more months.

Dependency Status. If a student is considered financially dependent, the parents' assets and income are considered in determining aid eligibility. If the student is financially independent, only the student's assets are considered, regardless of the relationship between student and parent. The specific definition of dependency status has varied across surveys. In the 1995–96 NPSAS, a student is considered independent if (1) the institution reports that the student is independent, or (2) the student meets one of the following criteria: (a) is age 24 or older at the end of the fall term of the NPSAS year; (b) is a veteran of the U.S. Armed Forces; (c) is an orphan or ward of the court; (d) is enrolled in a graduate or professional program beyond a bachelor's degree; (e) is married; (f) has legal dependents other than spouse.

Expected Family Contribution (EFC). The amount of financial support for the student's undergraduate education that is expected to be provided by the student's family, or directly by the student if the student is financially independent. This amount is used to determine financial need and is based upon dependency status (see above definition), family income and assets, family size, and the number of children enrolled in postsecondary education. If this information is not available from the institution, it is gathered from the Department of Education's financial aid system (the Central Processing System, or CPS) or it is imputed from student income.

Title IV Financial Aid. Sum of the following types of federal aid: Pell Grants, Supplemental Educational Opportunity Grants, Perkins Loans, Stafford Loans, PLUS Loans, and Federal Work Study.

4. SURVEY DESIGN

Target Population

The *survey population* is defined as those students who are enrolled in any term that begins between May 1 of one year and April 30 of the next year, thus allowing the student lists needed for sample selection to be obtained in January or February for most institutions. This definition was used starting with the 1992–93 NPSAS, and provides substantial comparability with the survey populations for the 1986–87 and 1989–90 NPSAS studies. Nearly all members of the target population are also members of the survey population. The population includes both students who receive aid and those who do not receive aid. It excludes students who are enrolled solely in a GED program or are concurrently enrolled in high school.

To be eligible for inclusion in the NPSAS institutional sample, an institution must satisfy the following conditions: (1) offer an education program designed for persons who have completed secondary education; (2) offer an academic, occupational, or vocational program of study lasting at least 3 months or 300 clock hours; (3) offer courses to the general public; (4) offer more than just correspondence courses; (5) be located in the 50 states, the District of Columbia, or Puerto Rico; (6) be other than a U.S. Service Academy.

Full-time and part-time students enrolled in academic or vocational courses or programs at these institutions, and not concurrently enrolled in a high school completion program, are eligible for inclusion in NPSAS.

Sample Design

The design for the NPSAS sample involves the selection of a nationally representative sample of postsecondary education institutions and students within those institutions. Prior to the 1995–96 study, NPSAS used a geographic-area-clustered, three-stage sampling design: (1) constructing geographic areas from three-digit postal zip code areas; (2) sampling institutions within the geographic sample areas; and (3) sampling students within sample institutions. The 1995–96 sample design eliminated the first stage of sampling (geographic area), thereby increasing the precision of the estimates. Over 950 postsecondary institutions, 50,000 students, and 8,800 parents were selected for participation in the 1995–96 NPSAS.

Institution sample. The institution-level sampling frame is constructed from the Integrated Postsecondary Education Data Systems (IPEDS) Institutional Characteristics (IC) file—see chapter 14. Although the institutional sampling strata have varied across NPSAS administrations, in all years the strata have been formed by classifying institutions according to control (public or private) and level (length of program and highest degree offering). A stratified sample of institutions is then selected with probabilities proportional to size (pps). School enrollment, as reported in the IPEDS, defines the measure of size; enrollment is imputed if missing in the IPEDS file. Institutions with expected frequencies of selection greater than unity are selected with certainty. The remainder of the institutional sample is selected from the other institutions within each stratum. Additional implicit stratification is accomplished within each institutional stratum by sorting the stratum sampling frame in a serpentine manner by: (a) institutional level of offering; (b) the IPEDS IC-listed Bureau of Economic Analysis of the U.S. Department of Commerce Region; and (c) the institution measure of size. This allows the approximation of proportional representation of institutions on these measures. Selected institutions are requested to verify the IPEDS classification (institutional control and highest level of offering) and the calendar system used (including dates that terms started).

As noted above, the 1995–96 NPSAS was the first to employ a single-stage institutional sampling design, no longer constructing geographic areas as the initial step. The sampling frame was the 1993–94 IPEDS IC file; 9,468 of the 10,651 institutions on the file were deemed eligible for the 1995–96 NPSAS. The eligible institutions were stratified into nine strata based on institutional control and highest level of offering.

For the 1995–96 study, 973 institutions were selected—131 with certainty and the remaining 842 probabilistically. A total of 73 (7.5 percent) of the selected institutions were subsequently found to be ineligible. Eligibility varied considerably with level of offering and control, being markedly lower for less than 2-year institutions and private for-profit institutions. However, these differences were expected and were directionally consistent with results from prior NPSAS studies.

Student sample. The sampled institutions are requested to provide student enrollment lists with the following information on each student: full name, identification number, Social Security Number, and educational level (and in the 1995–96 NPSAS, an indication of first-time beginning student (FTB) status). The student sample is drawn from these lists (provided by 836 of the 900 eligible institutions in the 1995–96 NPSAS). The 1986–87 NPSAS sampled only those students enrolled in the fall of 1986. Beginning with the 1989–90 NPSAS, students enrolled at any time during the year have been eligible for the study. This design change provides the data necessary to estimate full-year financial aid awards.

Basic student sample. Students are sampled on a flow basis (using stratified systematic sampling) from the lists provided by the institutions. Steps are taken to eliminate both within-institution and cross-institution duplication of students. NPSAS classifies students by educational level as undergraduate, graduate, or first-professional students. The 1995–96 NPSAS further stratified undergraduate students as (1) potential first-time, beginning students (FTBs) and (2) other undergraduates. The FTBs make up the second cohort of the Beginning Postsecondary Students Longitudinal Study. (See chapter 17.) For the purpose of defining the first cohort of the Baccalaureate and Beyond Longitudinal Study (see chapter 18), the 1992–93 NPSAS broke down undergraduates into: (1) business major baccalaureate recipients, (2) other baccalaureate recipients, and (3) other undergraduates.

The student sample is allocated to the combined institutional and student strata (e.g., graduate students in public, 4-year, doctorate institutions). Initial student sampling rates are calculated for each sample institution using refined overall rates to approximate equal probabilities of selection within the institution-by-student sampling strata. These rates are sometimes modified to ensure that the desired student sample sizes are achieved.

In the 1995–96 NPSAS, adjustments to the initial sampling rates resulted in some additional variability in the student sampling rates and, hence, in some increase in survey design effects. However, these rate adjustment procedures were generally effective. The overall sample yield in the 1995–96 NPSAS was actually greater than expected (63,616 students vs. the target of 59,509). The student sample consisted of 23,612 FTBs; 27,536 other undergraduates; 9,689 graduate students; and 2,779 first-professional students. (See "Longitudinal samples" below for more detail on the sampling of FTBs.)

Student interview sample. Prior to collection of data from the students themselves, information is abstracted from institutional records for the sampled students. Students for whom no record abstracts are available or who are found to be ineligible during record abstraction are excluded from the interview data collection. Due to budget limitations, the 1995–96 NPSAS attempted computer-assisted telephone interviewing (CATI) for only a subsample of the basic student sample. These sampling procedures resulted in 51,195 students selected for Phase 1 of the 1995–96 CATI interviewing. A sample of nonrespondents to Phase 1 was selected for Phase 2 with specified rates based on the outcome of the Phase 1 efforts and the seven sampling strata; 25,766 students were selected for Phase 2.

Parent interview subsample. Of the students selected for the student interview, a subsample is selected for interviewing of their parents. In the Phase 1 CATI subsample of the 1995–96 NPSAS, students were designated for parent interviewing if they met one of the following criteria: they were dependent undergraduate students not receiving federal aid; they were dependent undergraduate students receiving federal aid, whose parents' adjusted gross income was not available; or they were independent undergraduate students who were 24 or 25 years old on December 31, 1995. All 8,803 students who fell into one of these groups were sampled for parent interviews.

Longitudinal samples. In the 1989–90 NPSAS, a new longitudinal component collected baseline data for students who started their postsecondary education during 1989–90. These students are followed over time in the Beginning Postsecondary Students (BPS) Longitudinal Study. (See chapter 17.) Beginning postsecondary students from NPSAS 1995-96 were followed in 1998. Similarly, the 1992–93 NPSAS provided baseline data for students who received baccalaureates during the 1992–93 year. These graduates are followed over time as part of the Baccalaureate and Beyond (B&B) Longitudinal Study. (See chapter 18.)

Full-time Beginning (FTB) sample. Prior to the 1995–96 NPSAS, a ***pure FTB*** was defined as a student who enrolled in postsecondary education for the first time after high school during the NPSAS year. This definition was refined for the 1995–96 NPSAS to include students who had previously enrolled but ***had not completed a postsecondary course for credit*** prior to July 1, 1995 (referred to as *effective FTBs*). This expanded definition shifted the requirement from the act of enrollment to successful completion of a postsecondary course.

FTB status was determined in three stages—during student list acquisition, CADE institutional record abstraction, and CATI interviewing.

First, FTBs were sampled from the student lists provided by the institutions. However, information available to institutions was often insufficient for determining an accurate count of FTBs; for example, students transferring from another institution without transfer credits might mistakenly have been counted as FTBs. FTB sampling rates in the 1995–96 NPSAS were based primarily on the field test results and the previous BPS experience in the 1989–90 NPSAS, which indicated that the number of students listed as potential FTBs who were not actual FTBs far exceeded the number of students not identified as potential FTBs who later proved to be FTBs. As in the past, the 1995–96 NPSAS longitudinal cohort was oversampled to support the next BPS survey.

The second stage of FTB determination involved the screening of FTB status during abstraction of institutional records. Students classified as undergraduates were identified as potential FTBs for CATI subsampling based on year of high school graduation, birth year, and year-in-school variables. In the third and last stage, a number of FTB-screening questions in the student CATI interview allowed final determination of FTB status.

Baccalaureate sample. Baccalaureate recipients were classified as business major or other major. Some of the students on the graduation lists provided by the sample institutions were not actually scheduled to receive their baccalaureate degrees during the defined NPSAS year.

Data Collection and Processing

NPSAS relies on an integrated system of computer assisted data capture approaches: (a) electronic data interchange (EDI) with extant government databases, (b) computer-assisted data entry (CADE) of student financial aid records at institutions, and (c) computer-assisted telephone interviewing (CATI) of students and parents. Participating institutions designate Institutional Coordinators through which all communications are directed, including the provision of student enrollment lists for student sampling.

Reference dates. Data are collected for the financial aid award year, which spans from July 1 of one year through June 30 of the following year.

Data collection. NPSAS involves a multistage effort to collect information related to student aid. The 1995–96 study was the first to include an initial stage where Student Aid Report information from the Department of Education Central Processing System for federal aid applications was directly collected through EDI.

The second stage of data collection involves abstracting information from the student's records at the school from which he or she was sampled. Starting with the 1992–93 NPSAS, these data have been collected through a CADE system, which facilitates both collection and transfer of the information to subsequent electronic systems. To reduce respondent burden, several data elements are preloaded into CADE records prior to collection at the institution. These include student demographics, Student Aid Report information on federal financial aid applicants, and nonfederal aid common to a particular institution. Institutional Coordinators are given the option of having their staff or contractor field data collectors perform the data abstractions (guided by the CADE program). In the 1995–96 NPSAS, 57 percent of the institutions chose self-CADE.

In the third stage of data collection, information pertaining to family circumstances, background demographic data, and educational and work experiences and aspirations is obtained from students and a subsample of their parents. Student and parent questionnaires were used to collect this information in the first (1986–87) NPSAS. Beginning with the 1990–91 NPSAS, student and parent data have been collected by CATI. Unlike previous studies, the 1995–96 NPSAS interviewed only a subsample of students. Interviews were conducted in two phases, with potential first-time beginning students (FTBs) and federal aid applicants selected with certainty for Phase 1.

The principal form for the student interview contains 10 sections and is programmed for CATI administration. There are also three types of abbreviated interviews. One abbreviated form is for CATI administration to Spanish speakers with limited English proficiency. A second form is reproduced in Spanish and English language hardcopy for mailout to students who cannot be reached by phone, who indicate that they will only participate by mail, or who are hearing impaired (with eligibility established through Telephone Display for the Deaf). A third form is used for the reliability reinterview study, which is administered to a randomly selected subsample of students about 4 weeks after the full student interview. In addition, a minimal interview is used for CATI administration to sample members who have refused to participate on at least two different occasions, but who agree to answer a few questions in 5 minutes or less.

The parent supplement interview is maintained within the same record as the student interview (only in 1995–96), allowing the parent to be interviewed "on the spot" should that parent be contacted in attempting to locate the student.

Online coding is required for postsecondary education institution, major field of study, and industry/occupation. Institutions other than the sample institution are assigned their six-digit IPEDS identifier. Coding of major field of study and industry/occupation use a dictionary of word/code associations. When the interviewer enters the verbatim text provided by the respondent, standard descriptors associated with identified codes are displayed. The interviewer then selects one of the listed descriptors.

The final stage of data collection involves retrieval of additional Student Aid Report data (for the academic year beyond the NPSAS year) from the Central Processing System; data on Pell Grant applications for the NPSAS year from the Pell Grant file; and loan histories of applicants for federal student loans from the NSLDS (National Student Loan Data System). All of these files are maintained by the Department of Education.

Information has been collected on more than 55,000 students in every NPSAS administration.

Editing. Initial editing takes place during data entry. The CADE system has built-in quality control checks to notify the user of any student records that are incomplete (and the area of incompleteness) and any records that have not yet been accessed. A pop-up screen provides overall full and partial completion rates for institutional record abstraction. Once the contractor receives an institution's CADE package, every record is subjected to edit checks for completeness of critical items. Data from an institution fail the edit check if 50 percent or more of the student records fail all edit checks or if any anomalous data patterns are observed.

Following the completion of data collection, all CADE and CATI data are edited to ensure adherence to range and consistency checks. Range checks are summarized in the variable descriptions contained in the data files. Inconsistencies, either between or within data sources, are resolved in the construction of derived variables. The edit program also checks specific CATI items for validity by comparing the CATI responses to information available in institutional records. Missing data codes characterize blank fields as: don't know/data not available; refused; legitimate skip; data source not available (not applicable to the student); or other.

Estimation Methods

Weighting is used to adjust NPSAS data to national population totals and to adjust for unit nonresponse. Imputation is used to compensate for item nonresponse.

Weighting. For the purpose of obtaining nationally representative estimates, sample weights are created for both the institution and the student. Additional weighting adjustments, including nonresponse and poststratification adjustments, compensate for potential nonresponse bias and frame errors (differences between the survey population and the ideal target population). Multiplicity and trimming adjustments are also performed.

The 1995–96 NPSAS database contains a total of eight analysis weights associated with the CADE respondents, CATI respondents, and Study respondents. Weights are included for separate analyses on all students, undergraduate students, graduate students, and first-time beginning students (FTBs).

The CADE and CATI weights apply, respectively, to student respondents with CADE institutional record abstracts and CATI interviews. The Study weights apply to students who responded to specified CADE or CATI data items.

Study and CATI weights. The 1995–96 NPSAS Study weights and CATI weights were calculated as the product of 14 weight components, each representing either a probability of selection or a weight adjustment. Since the Study weights were restricted to students selected for CATI, the first nine weight components of the Study weights and CATI weights were identical; these represent the sample selection and adjustment components through the first phase of CATI. The remaining weight components followed the same steps, but calculations were performed separately because of the different response definitions.

FTB weights. FTBs whose first postsecondary institution was not the NPSAS sample institution were not to be included in the Beginning Postsecondary Students Longitudinal Study. To compensate for excluding these FTBs, the FTB weights were computed by making a final weighting class adjustment to the CATI weights by institution type. All adjustment factors were close to one, ranging from 1.00 to 1.02.

CADE weights. The development of the CADE weight components was similar to the development of the Study and CATI weight components—except that the CADE components applied to a different set of respondent data and did not include the CATI weight components.

Imputation. After the editing process (including logical imputations) is completed, the remaining missing values for several analysis variables (22 in the 1995–96 NPSAS) are statistically imputed in order to reduce the bias of survey estimates caused by missing data. Except for expected family contribution (EFC), which is imputed through a multiple regression approach, all variables are imputed using a weighted sequential hot deck procedure.

The respondent data for six key items are modeled using a Chi-squared Automatic Interaction Detector (CHAID) analysis to determine the imputation classes. These items are race/ethnicity, parent income (for dependent students only), student income, student marital status, dependents indicator, and number of dependents.

The other 15 items imputed by the weighted hot-deck approach in the 1995–96 NPSAS were: parent family size, parent marital status, student citizenship, student gender, student age, dependency status, local residence, type of high school degree, high school graduation year, fall enrollment indicator, attendance intensity in fall term, student level in last term, student level in first term, degree program in last term, and degree program in first term. Only four of these items had more than 5 percent of cases imputed: parent family size (18.0 percent); parent marital status (15.5 percent); high school degree (5.3 percent); and high school graduation year (5.3 percent).

As noted above, a regression approach is used to impute expected family contribution (EFC). The goal is to obtain the most parsimonious and best fitting equations using information likely to be available for nonaided students (those most likely to have a missing EFC). The general approach is to develop logistic regression models to estimate zero EFC cases, and then use ordinary least squares regression models to estimate the predicted EFC for nonzero EFC cases.

Recent Changes

The 1995–96 NPSAS included important new features in sample design and data collection. It was the first NPSAS to employ a single-stage institutional sampling design (no longer using an initial sample of geographic areas and institutions within geographic areas). This design change increased the precision of study estimates. The 1995–96 study was also the first NPSAS to select a *subsample* of students for telephone interviews, and to take full advantage of extant government data files. Through Electronic Data Interchange (EDI) with the Department of Education's Central Processing System, the study obtained financial data on federal aid applicants for both the NPSAS year and the year after. Through EDI with the National Student Loan Data System, full loan histories were obtained. Cost efficiencies were introduced through a dynamic two-phase sampling of students for computer-assisted telephone interviewing, and the quality of collected institutional data was improved through an enhanced CADE procedure. New procedures were also introduced to broaden the base of postsecondary student types for whom telephone interview data could be collected: the use of Telephone Display for the Deaf technology to facilitate telephone communications with hearing-impaired students, and a separate Spanish translation interview for administration to students with limited English language proficiency. In addition, students were oversampled to yield enough FTBs to serve as the second cohort for the Beginning Postsecondary Students Longitudinal Study.

Future Plans

The next round of surveys for NPSAS is scheduled for 2003–04; this survey will also serve as the start of another BPS longitudinal cohort.

5. DATA QUALITY AND COMPARABILITY

Every major component of the study is evaluated on an ongoing basis so that necessary changes can be made and assessed prior to task completion. Separate training is provided for CADE and CATI data collectors, and interviewers are monitored during CATI operations for deviations from item wording and skipping of questions. The CATI system includes online coding of postsecondary education institution, major field of study, and industry/occupation so that interviewers can request clarification or additional information at the time of the interview. Quality circle meetings of interviewers, monitors, and supervisors provide a forum to address work quality, identify problems, and share ideas for improving operations and study outcomes. Even with such efforts, however, NPSAS—like every survey—is subject to various types of errors, as described below.

Sampling Error

Because NPSAS samples are probability-based samples rather than simple random samples, simple random sample techniques for estimating sampling error cannot be applied to these data. Two common procedures for estimating variances of such survey statistics are the Taylor Series linearization procedure and the Jackknife replicate procedure, which are both available for use with NPSAS data.

Taylor Series. For the 1995–96 NPSAS, analysis strata and replicates for three separate data sets were defined: all students, all undergraduate students, and all graduate/first-professional students.

Jackknife. In the 1995–96 NPSAS, the Jackknife analysis strata were defined to be the same as the analysis strata defined for the Taylor Series procedure. Based on the Jackknife strata and replicate definitions, seven replicate weight sets were created—one set for the CADE weights and three sets each for the Study and CATI weights. The Study and CATI sets included separate replicate weights for all students, undergraduates only, and graduates only.

Nonsampling Error

Coverage error. Because the institutional sampling frame is constructed from the IPEDS IC file, there is nearly complete coverage of the institutions in the target population. Student coverage, however, is dependent upon enrollment lists provided by the institutions. In the 1995–96 NPSAS, 93 percent of the 900 eligible sample institutions provided student lists or databases that could be used for sample selection. As in prior NPSAS implementations, participation was highest among public institutions and lowest among private for-profit institutions.

Several checks for quality and completeness of student lists are made prior to actual student sampling. In the 1995–96 NPSAS, *completeness checks* failed if (1) FTBs were not identified (unless the institution explicitly indicated that no such students existed), or (2) student level (undergraduate, graduate, or first professional) was not clearly identified. *Quality checks* were performed by comparing the unduplicated counts (by student level) on institution lists with nonimputed unduplicated counts in IPEDS IC files. Institutions failing these checks were called to rectify the problems before sampling began. Almost half of the institutions provided lists with one or more problems. Well over one-third of the institutions had "suspect" counts, and more than one-tenth failed to identify FTBs.

Nonresponse error. The response rates described below refer to the 1995–96 NPSAS.

Unit nonresponse. There are several types of participation/coverage rates in NPSAS. For each type in the 1995–96 NPSAS, rates were generally lowest among for-profit institutions and institutions whose highest offering is less than a 4-year program.

In the 1995–96 NPSAS, 93 percent of eligible sample institutions provided student enrollment lists for student sampling. Of this group, 96 percent also provided full or partial CADE data from administrative records for at least one student (*institution CADE response rate*). The weighted and unweighted rates for institution CADE were quite comparable (90–100 percent), with a relatively small range of variation by institution type. The *student CADE coverage rate* was 93 percent (both unweighted and weighted). By institution type or student level, unweighted student coverage rates ranged from 88 to 96 percent, and weighted rates ranged from 81 to 97 percent.

For the subsample of students who were interviewed by telephone, the overall *student CATI response rate* was 76 percent weighted, with a range of 69 to 82 percent across domains (institutional type, student level, federal aid application status). Rates were uniformly higher for federal aid applicants than for nonapplicants. The *parent CATI* *response rate* for the parent subsample was 67 percent unweighted. This lower rate (as compared to student interviews) reflects the lower priority of parent interviews.

To determine the adequacy of coverage for analyses, an *overall study student yield rate* was computed, based on the following definition of a "yielding case": (1) the student CADE was effectively complete (Section 2 enrollment and tuition items were complete; the characteristics and subsection of Section 1 was complete; and either Section 3 was complete or comparable information was obtained from the Central Processing System, Pell Grant file, or the National Student Loan Data System), *or* (2) the Section A items of the student CATI were sufficiently complete to identify FTBs, or an abbreviated or minimal version of the student interview was completed. The overall study yield rate for the 1995–96 NPSAS was 97.0 percent unweighted and 96.3 percent weighted. Weighted and unweighted yield rates were quite consistent across domains (institutional type, student level), exceeding 92 percent in all cases.

The table below shows response rates across NPSAS administrations.

Table 6. Weighted response rates for selected NPSAS components

Component	List participation rate	Response rate	Overall
NPSAS 1989–90			
Student survey (analysis file)	86	84	72
Student survey (CATI resp.)	86	76	65
NPSAS 1992–93			
Student survey (analysis file)	88	75	66
Student survey (CATI resp.)	88	67	59
NPSAS 1995–96	*93	*81	*76

*Unweighted response rate
SOURCE: Seastrom, Salvucci, Walter, and Shelton (forthcoming), *A Review of the Use of Response Rates at NCES.*

Item nonresponse. Each NPSAS institution is unique with regard to the type of data maintained for its students. Because not all desired information is available at every institution, the CADE software allows entry of a "data not available" code. In the 1995–96 NPSAS, the percentage of missing responses was low for most CADE items, with only 12 items having nonresponse rates greater than 10 percent. More than half of these items pertained to undergraduate and graduate entrance examinations or higher institution degree. Four were demographic items: marital status, Hispanic ethnicity, race, and veteran status.

For student CATI interviews, item nonresponse rates were also fairly low. Only 54 of the more than 1,000 variables in the final CATI data set had more than 10 percent missing data (a combination of refusals and "don't knows"). Items with the largest amount of nonresponse pertained to undergraduate and graduate entrance examination scores; two-thirds or more of the students reporting that they had taken the SAT or GRE were unable to recall their scores. Questions most likely to evoke explicit refusals concerned student and parent income, assets, and debt; these also had high rates of "don't know."

Measurement error. Due to the complex design of NPSAS, there are several possible sources of measurement error, as described below.

Sources of response. Each source of information in NPSAS has both advantages and disadvantages. While students and their parents are more likely than institutions to have a comprehensive picture of education financing, they may not remember or have records of exact amounts and sources. This information may be more accurate in student financial aid records and government databases since it is recorded at the time of application for aid. Other information is likely to be most accurate when obtained from a parent; this is especially true for parents' finances.

Institutional records. While financial aid offices maintain accurate records of certain types of financial aid at that institution, these records are not necessarily inclusive of all support and assistance. They may not contain financial aid provided at other institutions attended by the student, and they may not include employee educational benefits and institutional assistantships, which are often treated as employee salaries. These amounts are assumed to be underreported.

Government databases. Federal aid information can only be extracted from federal financial aid databases if the institution can provide a valid Social Security Number for the student. It is likely that there is some undercoverage of federal aid data in NPSAS.

CATI question delivery. Any deviation from item wording that changes the intent of the question or obscures the question meaning can result in misinterpretation on the part of the interviewee and an inaccurate response. An interviewer's skipping of questions adds to the nonresponse rate. In the 1995–96 NPSAS, the cumulative question delivery error rate was less than 2 percent.

CATI data entry. CATI entry error occurs when the response to a question is recorded incorrectly. While these error rates were somewhat higher in the 1995–96 NPSAS than expected, problems were detected early and the CATI interviewers were retrained. Thus, the entry error rates show a consistent decline over the data collection period. The facility average error rate for the monitoring period was less than 2 percent.

Reinterview results. Reliability interviews are administered to a randomly selected subsample of students about 4 weeks after the full student interview. The reinterview questions broadly represent the student interview but are most heavily weighted to cover financial aid, financial support for educational expenses from family, educational status of family members, and student's work experiences while enrolled in the institution. Reliability indices for the educational finance items in the 1995–96 NPSAS were generally acceptable but somewhat mixed. While all items showed a more than 80 percent agreement between the interview and reinterview, the relational statistic only exceeded 0.80 for two items. In addition, two of the three items on work experience showed only marginally acceptable reliability (less than 70 percent), although the third item showed good reliability. All but one of the items related to personal and family educational experiences were reliable. The results for the income items were somewhat mixed.

Data Comparability

As noted in section 4, important design changes were implemented in the 1995–96 NPSAS. While sufficient comparability in survey design and instrument was maintained to ensure that comparisons with past NPSAS studies could be made, the data from the last three studies are not comparable to the first (1986–87) NPSAS for the following reasons: (1) the 1986–87 NPSAS only sampled students enrolled in fall 1986, whereas the later studies sampled from enrollments covering a full year; and (2) the 1986–87 NPSAS did not include students from Puerto Rico, whereas the studies since 1989–90 have included a small sample of Puerto Rican students. However, users of NPSAS data files can produce estimates for the later studies comparable to 1986–87 by selecting only students enrolled in the fall and excluding those sampled from Puerto Rico. Note also that the method used to generate the lists of students from which to sample was changed for the 1992–93 and subsequent NPSAS surveys.

Comparisons with IPEDS data. NCES recommends that readers not try to produce their own estimates (e.g., the percentage of all students receiving aid or the numbers of undergraduates enrolled in the fall who received federal aid, state aid, etc.) by combining estimates from NPSAS publications with the IPEDS enrollment numbers. The IPEDS enrollment data are for fall enrollment only and include some students not eligible for NPSAS (e.g., those enrolled in U.S. Service Academies and those taking college courses while enrolled in high school).

6. CONTACT INFORMATION

For content information on NPSAS, contact:

Aurora M. D'Amico
Phone: (202) 502–7334
E-mail: aurora.d'amico@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Methodology Report for the 1990 National Postsecondary Student Aid Study, NCES 92–080, by Westat, Inc. Washington, DC: 1992.

Methodology Report for the National Postsecondary Student Aid Study, 1987, NCES 90–309, by Westat, Inc. Washington, DC: 1990.

Methodology Report for the National Postsecondary Student Aid Study, 1992–93, NCES 95–211, by J.D. Loft, J.A. Riccobono, R.W. Whitmore, R.A. Fitzgerald, and L.K. Berkner. Washington, DC: 1995.

National Postsecondary Student Aid Study, 1995–96 (NPSAS:96) Methodology Report, NCES 98–073, by J.A. Riccobono, R.W. Whitmore, T.J. Gabel, M.A. Traccarella, D.J. Pratt, and L.K. Berkner. Washington, DC: 1997.

Survey Design

National Postsecondary Student Aid Study: 1996 Field Test Methodology Report, NCES Working Paper 96–17, by Research Triangle Institute. Washington, DC: 1996.

Data Quality and Comparability

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, and M. Saba. Washington, DC: 1997.

A Review of the Use of Response Rates at NCES (forthcoming), by M. Seastrom, S. Salvucci, E. Walter, and K. Shelton.

Chapter 17: Beginning Postsecondary Students (BPS) Longitudinal Study

1. OVERVIEW

The Beginning Postsecondary Students (BPS) Longitudinal Study was implemented in 1990 to complement the NCES longitudinal studies of high school cohorts and improve data on participants in postsecondary education. BPS draws its cohorts from the National Postsecondary Student Aid Study (NPSAS), an information system that regularly collects financial aid and other data on nationally representative cross-sectional samples of postsecondary students. (See chapter 16.) NPSAS provides the base year data for first-time beginning (FTB) postsecondary students; BPS then follows these students through school and into the workforce.

BPS includes nontraditional (older) students as well as traditional students and is, therefore, representative of *all* beginning students in postsecondary education. By starting with a cohort that has already entered postsecondary education and following it every 2–3 years for at least 6 years, BPS can describe to what extent, if any, students who start their education later differ in progress, persistence, and attainment from students who start earlier. In addition to the student data, BPS collects financial aid records covering the entire undergraduate period, providing complete information on progress and persistence in school.

The first BPS cohort identified about 8,000 first-time beginning students who began their postsecondary education in the 1989–90 academic year; this cohort was followed up in 1992 and 1994. The second BPS cohort, which followed about 10,200 students who started their postsecondary education in the 1995–96 academic year, was followed up in 1998 and 2001. A third BPS cohort is planned for 2003–04, in conjunction with that NPSAS data collection.

LONGITUDINAL SAMPLE SURVEY OF FIRST-TIME BEGINNING POSTSECONDARY STUDENTS, INCLUDING BOTH TRADITIONAL AND NONTRADITIONAL STUDENTS

BPS includes:

- Base year NPSAS data
- Student interviews
- Financial aid records

Purpose

To collect data related to persistence in and completion of postsecondary education programs; relationships between work and education; and the effect of postsecondary education on the lives of individuals.

Components

BPS consists of base year data obtained from NPSAS, follow-up data collected in BPS surveys, and student aid records from ED Pell grant and loan files.

Base Year Data (from NPSAS). Information includes data collected in NPSAS from students, parents, institutional records, and Department of Education financial aid records. This includes information such as: major field of study; type and control of institution; financial aid; cost of attendance; age; sex; race/ethnicity; family income; reasons for school selection; current marital status; employment and income; community service; background and preparation for college; college experience; future expectations; parents' level of education; income; and occupation. These data represent the 1989–90 academic year for the first BPS cohort and the 1995–96 academic year for the second cohort.

BPS Follow-up Surveys. Follow-up data are obtained from student interviews and financial aid records: year in school; persistence in enrollment; academic progress; degree attainment; change in field of study; institution transfer; education-related experiences; current family status; expenses and financial aid; employment and income; employment-related training; community service; political participation; and future expectations. BPS follows each cohort twice at 2–3 year intervals.

Periodicity

BPS cohorts are followed at least twice after first entering postsecondary education (as determined in NPSAS). Follow ups take place at 2–3 year intervals.

2. USES OF DATA

BPS addresses persistence, progress, and attainment after entry into postsecondary education and also directly addresses issues concerning entry into the workforce. Its unique contribution is the inclusion of nontraditional (or older) students—a steadily growing segment of the postsecondary student population. Their inclusion allows analysis of the differences, if any, between traditional (recent high school graduates) and nontraditional students in aspirations, progress, persistence, and attainment.

Congress and other policymakers use BPS data when they consider how new legislation will affect college students and others in postsecondary education. BPS data can answer such questions as: What percentage of beginning students complete their degree programs? What are the financial, family, and school-related factors that prevent students from completing their programs, and what can be done to help them? Do students receiving financial aid do as well as those who do not? Would it be better if the amount of financial aid was increased? Additional questions that BPS can address include: Do students who are part-time or discontinuous attenders have the same educational goals as full-time, consistent attenders? Are they as likely to attain similar educational goals? Are students who change majors more or less likely to persist?

3. KEY CONCEPTS

Some of the key concepts in BPS are defined below.

Institution Type. Defined by level of degree offering and length of program at the postsecondary institution. Institutions are generally classified as: less-than-2-year (offers only programs of study that are less than 2 years in duration); 2- to 3-year, sometimes referred to in reports as 2-year (confers at least a 2-year formal award but not a baccalaureate, or offers a 2- or 3-year program that partially fulfills requirements for a baccalaureate or higher degree at a 4-year institution; includes most community and junior colleges); and 4-year (confers at least a baccalaureate degree and may also confer higher level degrees, such as master's, doctoral, and first-professional degrees; this category is often broken down into doctorate-granting vs. nondoctorate-granting).

Institution Control. Control of postsecondary institution, classified as follows: (1) public; (2) private, not-for-profit; and (3) private, for-profit.

First-time Beginning Students (FTBs). The target population for BPS. For the first BPS cohort, FTBs were defined as students who enrolled in postsecondary education for the first time after high school in the 1989–90 academic year (*pure FTBs*). Individuals who started postsecondary education earlier, left, and then returned were not included. The second BPS cohort comprised both students who enrolled for the very first time in the 1995–96 academic year *and* students who had previously enrolled but *had not completed a postsecondary course for credit* prior to July 1, 1995 (*effective FTBs*). This expanded definition shifted the requirement from the act of enrollment to successful completion of a postsecondary course.

Nontraditional Students. Primarily older students who delayed postsecondary enrollment; that is, did not enter postsecondary education in the same calendar year as high school graduation or received a general equivalency diploma (GED) or other certificate of high school completion.

Persistence. Continuous enrollment in postsecondary education with the goal of obtaining a degree or other formal award.

Attainment. Receipt of the degree or other formal award that was the student's objective while enrolled in postsecondary institutions.

Socioeconomic Status (SES). A composite variable combining parents' educational attainment and occupational status, dependent student's family income, and the existence of a series of material possessions in the respondent's home.

4. SURVEY DESIGN

Target Population

All students who first entered postsecondary education after high school in the 1989–90 academic year (the first BPS cohort) or in the 1995–96 academic year (the second BPS cohort). The definition of a first-time beginning student (FTB) was refined for the second BPS cohort to include students who had enrolled in postsecondary education prior to completion of high school as long as they had not completed a postsecondary course for credit before July 1, 1995 (the beginning of the 1995–96 academic year). BPS includes students in nearly all types of postsecondary education institutions located in the 50 states, the District of Columbia, and Puerto Rico: public, private not-for-profit, and private for-profit institutions; 2-year, 2- to 3-year, and 4-year institutions; and occupational programs that last for less than 2 years. Excluded are students attending U.S. Service Academies, institutions that offer only correspondence courses, or institutions that enroll only their own employees. BPS data are nationally representative by institutional level and control; the data are not representative at the state level.

Sample Design

Student eligibility for BPS is determined in two stages. The first stage involves selection for the base year NPSAS sample (the 1989–90 NPSAS for the first BPS cohort; the 1995–96 NPSAS for the second BPS cohort); see chapter 16 for a description of NPSAS sample design and determination of first-time beginning students (FTBs) who make up the BPS cohorts. All FTBs who complete interviews in NPSAS are considered eligible for BPS. The second stage of FTB determination involves a review of NPSAS data to see if any potential FTBs have been misclassified. FTB status for additional students may be determined through: (1) reports from NPSAS institutions; (2) responses of the sample member during the BPS interview; and (3) modeling procedures used following data collection.

First BPS cohort (1989–90). The first BPS cohort initially consisted of 11,700 students (from 1,092 institutions) who had been interviewed in the 1989–90 NPSAS. In the second follow up of this cohort in 1994, a working sample of 7,914 individuals was initially used. It consisted of the first follow-up eligible respondents, plus those nonrespondents for whom FTB status had yet to be determined. Only 7,132 sample members could be located. Of these, 6,786 members were interviewed, either fully or partially. Some of those interviewed (169) were determined to be non-FTBs, leaving 6,617 eligible FTBs who were either fully (5,926) or partially (691) interviewed in the second follow up.

Second BPS cohort (1995–96). In the second BPS cohort, 12,410 confirmed and potential FTBs were selected (from 788 institutions) for continued follow up from a total NPSAS pool of 15,728 confirmed or potential FTBs. This pool included 3,743 who had not been interviewed in the 1995–96 NPSAS (of which 425 were selected for potential continued inclusion in BPS). This BPS-eligible sample of 12,410 individuals was further reduced when an additional 230 were determined to be ineligible. The final BPS-eligible sample contained 10,268 FTBs who were given full or partial interviews in the first follow up; 1,060 were not able to be contacted, and 852 did not respond.

The final sample for this cohort includes 10,367 individuals. This includes all respondents to earlier follow ups as well as a subsample of earlier nonrespondents and other individuals who were unavailable for earlier data collections.

Data Collection and Processing

Computer-assisted telephone interviewing (CATI) is the primary data collection tool in BPS. All locating, interviewing, and data processing activities are under the control of an Integrated Control System (ICS), consisting of a series of PC-based, fully linked modules. The various modules of the ICS provide the means to conduct, control, coordinate, and monitor the several complex, interrelated activities required in the study and to serve as a centralized, easily accessible repository for project data and documents. BPS is conducted for NCES by the Research Triangle Institute.

The following sections describe the procedures for BPS follow ups. Refer to chapter 16 for a description of data collection and processing for the base year data obtained from NPSAS.

Reference dates. The base year (NPSAS) survey largely refers to experiences in postsecondary schooling in the academic year covered by NPSAS (1989–90 for the first

BPS cohort; 1995–96 for the second BPS cohort). The follow ups cover the 2- to 3-year interval since the previous round of data collection. Some data are collected retrospectively for the previous survey.

Data collection. Data collection in BPS follow ups involves concerted mail and telephone efforts to trace potential sample members to their current location and to conduct a CATI interview both to establish study eligibility and collect data. Field location and computer-assisted personal interviewing (CAPI) were also used extensively with the second cohort.

Locating students begins with information provided by the BPS locating database, which is updated by a national change of address service before the locating effort. Cases not located during the previous round of the survey are forwarded to pre-CATI telephone tracing, and subsequently to field locating if intensive telephone tracing is unsuccessful. Prior to the start of CATI operations, a prenotification mailing is sent to the student, and the current contact information is provided to interviewers for basic CATI locating. In the event that CATI locating is unsuccessful, cases are sent to post-CATI central trace for telephone tracing and, again as necessary, field locating. During tracing operations, cases of "exclusion" are identified, such as those who are: (1) outside of the calling area; (2) deceased; (3) institutionalized or physically/mentally incapacitated and unable to respond to the survey; or (4) otherwise unavailable for the entire data collection period.

Throughout the data collection period, interviewers are monitored for delivery of questionnaire text and recognition statements, probing, feedback, and CATI entry errors.

Each coding operation is subjected to quality control review and recoding procedures by expert coders. Subsequent to data collection, all "other, specify" responses are evaluated for possible manual recoding into existing categories, or into new categories created to accommodate responses of high frequency through a process known as "upcoding." Efforts are also made to convert several items with high rates of undetermined response (including refusal or "don't know"). In order to reduce indeterminacy rates for personal, parent, and household income items, as well as for other financial amount items, specific questions are included in the survey to route initial "don't know" responses through a series of screens seeking closer and closer estimates for the financial questions. In the second follow up of the first BPS cohort, amount ranges for the "don't know" conversion screens were based on frequencies obtained from the second follow-up field test for the same items. Indeterminacy conversion was attempted for five financial amount items (financial aid amount, total loan amount, respondent gross income, parents' gross income, and household gross income) and was very successful for initial "don't know" responses. Conversion rates were greater than 50 percent for every item attempted, with an overall success rate of 65 percent.

Editing. The CATI data are edited and cleaned as part of the preparation of the data file. Modifications to the data are made, to the extent possible, based on problem sheets submitted by interviewers which detail item corrections, deletions, and prior omissions. In addition, variables are checked for legitimate ranges and interim consistency. Coding corrections and school information from the IPEDS IC files (see above) are merged into the CATI files. Data inconsistencies identified during analyses are also corrected, as appropriate and feasible.

Estimation Methods

Weighting is used to adjust for unit nonresponse. Only minimal imputation is performed to compensate for item nonresponse.

Weighting. BPS follow ups involve further identification of FTB status for sample members who were in the earlier round of BPS. Further, post hoc modeling is implemented following the first follow-up data collection in an attempt to identify non-FTBs among nonrespondents.

Four sets of weights were computed for use with BPS data for the first (1989–90) cohort: (1) 1992 cross-sectional weights for cross-sectional analyses of the first cohort at the time of the first follow up, based on the first follow-up data collection; (2) 1994 cross-sectional weights for cross-sectional analyses of the first cohort at the time of the second follow-up data collection; (3) 1992 cross-sectional weights for the first follow up information which was collected either during the first follow up or retrospectively in the second follow up; and (4) longitudinal weights for comparison of the responses pertaining to the 1990, 1992, and 1994 cross-sectional populations (e.g., trend analyses), for those students who responded to each of the three surveys: the 1989–90 NPSAS, the BPS first follow up in 1992, and the BPS second follow up in 1994. For computation of these weights, see the technical report for the second follow up.

The 1994 cross-sectional weights can also be used for longitudinal analyses involving data items collected retrospectively in the second follow up because those data items are available for 1992, either directly from the first follow up or retrospectively from the second follow up if the student responded in 1994. Each set of weights consists of an analysis weight for computing point estimates of population parameters, plus a set of 35 replicate weights for computation of sampling variances using the Jackknife replication method of variance estimation. All weight adjustments were implemented independently for each set of replicate weights. (See section 5, Sampling Error, for further detail on replicate variance estimation.)

Imputation. Imputation is performed on a small number of variables in BPS. These variables relate to the student's dependency status and family income in each survey round. For example, the variable containing dependency status for aid in academic year 1989–90 was derived by examining all applicable variables used in the federal definition of dependency for the purpose of applying for financial aid. If information was not available for all variables, dependency status was imputed based on age, marital status, and graduate enrollment. Similarly, the variable containing the 1988 family adjusted gross income used imputed values if responses were not available.

Future Plans

The second BPS cohort (1995–96 FTBs) was followed up for the first time in 1998; a second follow up took place in 2001. A third BPS cohort is planned for 2003–04, in conjunction with a new round of NPSAS data collection.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Because the NPSAS sample design involves stratification, disproportionate sampling of certain strata, and clustered (i.e., multistage) probability sampling, the standard errors, design effects, and the related percentage distributions for a number of key variables in BPS have been calculated with the software package SUDAAN. These variables include: sex, race/ethnicity, age in the base year, socioeconomic status, income/dependency in the base year, number of risk factors in the base year, level and control of the first institution, and aid package at the first institution in the base year. These estimates provide an approximate characterization of the precision with which BPS survey statistics can be estimated.

Several specific procedures are available for calculating precise estimates of sampling errors for complex samples. Taylor Series approximations, Jackknife repeated replications, and balanced repeated replications produce similar results.

Nonsampling Error

Nonsampling error in BPS is largely related to nonresponse bias caused by unit and item nonresponse and to measurement error.

Coverage error. The BPS sample is drawn from NPSAS. Consequently, any coverage error in the NPSAS sample will be reflected in BPS. (Refer to chapter 16 for coverage issues in NPSAS.)

Nonresponse error. Unit nonresponse is reported in BPS in terms of *contact rates* (the proportion of sample members who were located for an interview) and *interview rates* (the proportion of sample members who fully or partially completed the interview). Item nonresponse has not been fully evaluated, although the numbers of nonrespondents are in the electronic codebook (ECB) on an item-by-item basis.

Unit nonresponse. The results for the second follow up of the first BPS cohort show a *contact rate* of 91.6 percent. The rate was substantially lower for individuals who did not respond to the first follow up (75.1 percent) than for those who did respond (95.1 percent). Contact rates also varied by institutions. The rate was highest for sample members who attended 4-year colleges (95.1 percent); in contrast, contact was made with only 80.8 percent of sample members attending private for-profit institutions with programs of less than 2 years.

Among those students who were contacted for the second follow up, the *interview rate* was 95.2 percent. The rate was higher for respondents to the first follow up than for nonrespondents by almost 8 percentage points (96.3 percent vs. 88.6 percent, respectively). Interview rates were fairly similar across institutions—ranging from 90.5 percent for students attending less than 2-year private not-for-profit institutions to 96.0 percent for students attending 4-year private not-for-profit colleges.

The table below summarizes the unit level and overall level weighted response rates across BPS administrations.

Table 7. Unit level and overall level weighted response rates for selected BPS surveys

Survey	Unit level weighted response rates			
	Base year 1st level	Base year 2nd level	1st wave	2nd wave
Students	86	84	*82	91
	Overall level weighted response rates			
	Base year 1st level	Base year 2nd level	1st wave	2nd wave
Students	86	72	*71	78

*Unweighted response rate
SOURCE: Seastrom, Salvucci, Walter, and Shelton (forthcoming), *A Review of the Use of Response Rates at NCES.*

Item nonresponse. Overall item nonresponse rates have been low across surveys (only 10 of the 363 items in BPS:96/98 contained over 10 percent missing data). Items with the highest rates of nonresponse were those pertaining to income. Many respondents were reluctant to provide information about personal and family finances and, among those who are not, many simply do not know this information.

Measurement error. While comprehensive psychometric evaluations of BPS data have not been conducted, issues of data quality are addressed during data collection.

Cross-interview data verification. During data collection, information from a prior interview (or from base year NPSAS data) is verified or updated to ensure compatibility across survey waves. In the first follow up of the first BPS cohort, demographic information covered in NPSAS (e.g., sex, race, and ethnicity) was verified or updated. The results indicated high reliability of these items. Prior to the full-scale second follow up, another set of items covered in earlier rounds was verified or updated, including high school graduation status, schools attended prior to the base year, and jobs held prior to the base year. These data were also found to be reliable across survey waves. Agreement approached 100 percent on high school graduation status, 99 percent on previous attendance of postsecondary schools, and 96 percent on previous jobs.

Reinterview. All BPS interview activities have involved a reinterview of a subsample of respondents to the main interview for the purpose of evaluating consistency of responses to the two interviews. The interval between the initial interview and the reinterview was 7–14 weeks.

Across BPS data collections, each new reinterview is designed to build on previous analyses by targeting revised items, new items, and items not previously evaluated. The second follow-up reinterview design and analysis focused on items that were revised in the full-scale study questionnaire based on first follow-up field test reinterview results. Reinterview analyses focused on data items that were expected to be stable for the time period between the initial interview and the reinterview. These items covered education experience; work experience (e.g., employee primary role, future career plans, principal job's relation to education, satisfaction with principal job, and factors affecting employment goals); education finances; and living arrangements.

Reliability, as measured by rates of agreement between the two interviews, showed considerable variation. Items on education experience had relatively high rates of agreement between interviews, ranging from 86.6 to 96.6 percent. Items on work experience and its relation to postsecondary school and future plans had moderate agreement, ranging from 66.7 to 95.8 percent. The greatest variation was for the items on principal job in relation to education; agreement between the two interviews ranged from 42.1 to 90.3 percent. The reliability of measures of satisfaction with the most recent job, employment goals, and education finances was moderate, ranging from 63 to 96 percent. Items about living arrangements showed the highest agreement, with several items reaching 100 percent.

Item order effects. The second follow up of the first BPS cohort also included a field test of the item order effects, that is, the sequence in which questionnaire items are presented to the respondents and the resulting response patterns. Discrepancies were examined and adjustments were made, as required, in the full-scale data collection. Also analyzed were discrepancies of online coding procedures for postsecondary institutions, fields of study, and combined and separate industry and occupations. To achieve high data quality, expert coding personnel recoded items that had been identified as inconsistent.

6. CONTACT INFORMATION

For contact information on BPS, contact:

Aurora M. D'Amico
Phone: (202) 502–7334
E-mail: aurora.d'amico@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:90/92) Final Public Technical Report, NCES 94–369, by G.J. Burkheimer, Jr., B.H. Forsyth, R.W. Whitmore, J.S. Wine, K.M. Blackwell, K.J. Veith, and G.D. Borman. Washington, DC: 1994.

Beginning Postsecondary Students Longitudinal Study Second Follow-up (BPS:90/94) Final Technical Report, NCES 96–153, by D.J. Pratt, R.W. Whitmore, J.S. Wine, K.M. Blackwell, B.H. Forsyth, T.K. Smith, E.A. Becker, K.J. Veith, M. Mitchell, and G.D. Borman. Washington, DC: 1996.

Beginning Postsecondary Students Longitudinal Study First Follow-up 1996–98 (BPS:96/98) Methodology Report, NCES 2000–157, by J.S. Wine, R. W. Whitmore, R.E. Heuer, M. Biber, and D.J. Pratt. Washington, DC: 2000.

Survey Design

Beginning Postsecondary Students Longitudinal Study Field Test Methodology Report (BPS:90/92), NCES 92–160, by G.J. Burkheimer, Jr., B.H. Forsyth, S.C. Wheeless, K.A. Mowbray, L.M. Boehnlein, S.M. Knight, and K.J. Veith. Washington, DC: 1992.

Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96/98) Field Test Report, NCES Working Paper 98–11, by D.J. Pratt, J.S. Wine, R.E. Heuer, R.W. Whitmore, J.E. Kelly, J.M. Doherty, J.B. Simpson, and M.C. Norman. Washington, DC: 1998.

Data Quality and Comparability

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, and M. Saba. Washington, DC: 1997.

A Review of the Use of Response Rates at NCES (forthcoming), by M. Seastrom, S. Salvucci, E. Walter, and K. Shelton.

Chapter 18: Baccalaureate and Beyond (B&B) Longitudinal Study

1. OVERVIEW

The Baccalaureate and Beyond (B&B) Longitudinal Study provides information concerning education and work experiences following completion of the bachelor's degree. It provides both cross-sectional profiles of bachelor's degree recipients 1 year after degree award and longitudinal data concerning their entry into and progress through graduate level education and the workforce. Special emphasis is placed on those graduates entering public service areas, particularly teaching, and provides information on their entry into the job market and career path.

B&B draws the base year data for its cohorts from the National Postsecondary Student Aid Study (NPSAS, see chapter 16). The first B&B cohort consists of individuals who received a bachelor's degree in the 1992–93 academic year; a second cohort was formed from baccalaureate recipients in the 1999–2000 academic year, and went to the field in 2001. B&B expands the efforts of the former Recent College Graduates Survey to provide unique information on educational and employment-related experiences of these degree recipients over a longer period of time. The 1993 cohort will be followed several times over a 12-year period so that most respondents who attend graduate or professional schools will have completed (or nearly completed) their education and be established in their careers. B&B can address issues concerning delayed entry into graduate school, progress and completion of graduate level education, and the impact of undergraduate and graduate debt on choices related to career and family.

LONGITUDINAL SAMPLE SURVEY OF BACHELOR'S DEGREE RECIPIENTS; THREE FOLLOW UPS OVER A 10-YEAR PERIOD

B&B collects data from:

- Base Year NPSAS Data
- Student interviews
- Undergraduate transcripts
- Federal financial aid and loan records
- Identified newly qualified teachers

Purpose

To (1) provide information on college graduates' entry into, persistence and progress through, and completion of graduate level education in the years following receipt of the bachelor's degree; and (2) provide information on the career paths of new teachers: retention, defection, delayed entry, and movement within the educational system.

Components

B&B consists of base year data culled from NPSAS. NPSAS data are collected in three components: the Student Record Abstract, the Student Interview, and the Parent Interview. The first B&B follow-up survey in 1994 collected data from a Student Interview as well as from college transcripts for their undergraduate program. The second follow up, conducted in 1997, combined a Student Interview with Department Aid Application/Loan Records data. A second B&B cohort, consisting of 1999–2000 baccalaureate recipients, went to the field in 2001.

Base Year Data (from NPSAS). B&B obtains its base year information from NPSAS. The NPSAS Student Record Abstracts (institutional records) provide major field of study; type and control of institution; attendance status; tuition and fees; admission test scores; financial aid awards; cost of attendance; student budget information and expected family contribution for aided students; grade point average; age; and date first enrolled. The base year data also include information from NPSAS Student Interviews regarding educational level; major field of study; financial aid at other schools attended during the year; other sources of financial support; monthly expenses; reasons for selecting the school attended; current marital status; age; race/ethnicity; sex; highest degree expected; employment and income; community service; expectations for employment after graduation; expectations for graduate school; and plans to enter the teaching profession. Data taken from the NPSAS Parent Interviews include: marital status; age; highest level of education achieved; income; amount of financial support provided to children; types of financing used to pay child's educational expenses; and current employment (including occupation and industry).

B&B First Follow-up Survey. The first follow up is conducted 1 year after the bachelor's degree was received (e.g., 1994 for the 1992–93 B&B cohort). In the Student Interview portion of the survey, recent graduates provide information regarding employment after degree completion; job search activities; expectations for and entry into teaching; teacher certification status; job training and responsibilities; expectations/entry into graduate school; enrollment after degree; financial aid; loan repayment/status; income; family formation and responsibilities; and participation in community service. This is the only follow up planned for the 2000 cohort (in 2001). As part of the first follow up of the 1992–93 B&B cohort, the Undergraduate Transcript Study component collected transcripts providing the following information: undergraduate coursework; institutions attended; grades; credits attempted and earned; and academic honors earned. All transcript information is as reported by the institutions, converted to semester credits and a 4.0 grade scale for comparability.

B&B Second Follow-up Survey. The second follow up for the 1992–93 B&B cohort was conducted 4 years after the bachelor's degree was received, in 1997. Participants provided information in the Student Interview regarding their employment history; enrollment history; job search strategies at degree completion; career progress; current status in graduate school; nonfederal aid received; additional job training; entry into/persistence in/resignation from teaching career; teacher certification status; teacher career path; income; family formation and responsibilities; and participation in community service.

The second follow up of the 1992–93 B&B cohort also included a Department Aid Application/Loan Records component to collect information on the types and amounts of federal financial aid received, total federal debt accrued, and students' loan repayment status. One of the goals of B&B is to understand the effect education-related debt has on graduates' choices concerning their careers and further schooling.

B&B Additional Follow-up Surveys. The 1993 cohort will be followed for a third time in 2003. The 2000 cohort was followed only in 2001.

Periodicity

The two B&B cohorts each have their own follow-up schedule, as described above.

2. USES OF DATA

B&B covers many topics of interest to policymakers, educators, and researchers. For example, B&B allows analysis of the participation and progress of recent degree completers in the workforce, relationship of employment to degree, income and ability to repay debt, and willingness to enter public service-related fields. B&B also allows analysis of issues related to access and choice into graduate education programs. Here emphasis is on ability, ease, and timing of entrance into graduate school, and attendance/employment patterns, progress, and completion timing once entered.

The unique features of B&B allow it to be used to address issues related to undergraduate education as well as postbaccalaureate experiences. This information has been used to investigate the relationship between undergraduate debt burden and early labor force experiences, and between undergraduate academic experiences and entry into teaching. These and other relationships can be investigated both in the short term and over longer periods.

Because B&B places special emphasis on new teachers at the elementary and secondary levels, it can be used to address many issues related to teacher preparation, entry into the profession (e.g., timing, ease of entry), persistence in or defection from teaching, and career movement within the education system.

Major issues that B&B attempts to address include:

- Length of time following receipt of degree after which college graduates enter the workforce;
- Type of job which graduates obtain, compared with major field of undergraduate study;
- Length of time to complete degree;
- Length of time to obtain a job related to respondents' field of study;
- Extent to which jobs obtained relate to educational level attained by respondent;
- Extent to which level of debt incurred to pay for education influences decisions concerning graduate school, employment, and family formation;
- Extent to which level of debt incurred influences decisions to enter public service professions;
- Rates of graduate school enrollment, retention, and completion;
- Extent to which delaying graduate school enrollment influences respondent's access to and progression through advanced degree programs;
- Factors influencing the decision to enroll in graduate education;
- Extent to which attaining an advanced degree influences short-term and long-term earnings;
- Number of graduates qualified to teach;
- Extent to which degree level/profession influences rate of advancement; and
- Extent to which respondents change jobs or careers.

3. KEY CONCEPTS

Some of the concepts and terms used in the B&B data collection and analysis are defined below. For more information on these terms and others used in B&B, refer to *A Descriptive Summary of 1992–93 Bachelor's Degree Recipients 1 Year Later With an Essay on Time to Degree* (NCES 96–158).

Degree-granting Institution. Any institution offering an associate's, bachelor's, master's, doctor's, or first-professional degree. Institutions that grant only certificates or awards of any length (less than 2 years, or 2 years or more) are categorized as nondegree-granting institutions.

First Postsecondary Institution. The first institution attended by the respondent following high school and in which the respondent was enrolled for a minimum of 3 months. Institutions attended before high school graduation are included if enrollment continued after high school graduation. The first institution may or may not be the institution that granted the bachelor's degree.

Status in Teacher Pipeline. This variable measures extent of involvement with teaching, using variables from 1994 and 1997 interviews and composites. Respondents who taught were classified as having taught with certification, with student teaching, without training, or with training unknown. Those who did not teach were classified as certified, having student taught, applied for teaching jobs, considered teaching, or having no interest or action in teaching. An additional category of cases who had become certified but whose teaching status was unknown was identified. All of these categories were combined in various ways throughout the report, depending on the context of the particular analysis.

Dependency Level. If a student is considered financially dependent, the parents' assets and income are considered in determining aid eligibility. If the student is financially independent, only the student's assets are considered, regardless of the relationship between student and parent. The specific definition of dependency status has varied across surveys. In the 1995–96 NPSAS, a student is considered independent if (1) the institution reports that the student is independent, or (2) the student meets one of the following criteria: (a) is age 24 or older at the end of the fall term of the NPSAS year; (b) is a veteran of the U.S. Armed Forces; (c) is an orphan or ward of the court; (d) is enrolled in a graduate or professional program beyond a bachelor's degree; (e) is married; or (f) has legal dependents other than spouse.

4. SURVEY DESIGN

Target Population

All postsecondary students in the 50 states, the District of Columbia, and Puerto Rico who completed a bachelor's degree in the academic year 1992–93, spanning July 1, 1992 to June 30, 1993 (first B&B cohort) or in the academic year 1999–2000, spanning July 1, 1999 to June 30, 2000 (second B&B cohort). Students from United States Service Academies are excluded because they are not part of NPSAS, from which B&B draws its samples.

Sample Design

B&B cohorts are subsamples of the NPSAS samples. (See chapter 16 for description of the NPSAS sample design.) Students in a given NPSAS sample are considered potentially eligible for a given B&B cohort if there is information indicating that the student had received, or expected to receive, a baccalaureate degree in the NPSAS year (e.g., between July 1, 1992 and June 30, 1993 for the first B&B cohort). Eligibility is determined in two ways: first, by confirming with respondents the date they received their baccalaureate degrees, and second, by examining student transcripts received from baccalaureate institutions. All NPSAS sample persons who satisfy the subsample requirements are designated as eligible for the B&B sample irrespective of whether they were respondents or nonrespondents in NPSAS.

In order to provide a base year sample for the first B&B cohort (1992–93 bachelor's degree recipients), NCES introduced several design modifications into the 1992–93 NPSAS. First, the number of sample institutions offering only programs of less than 4 years was reduced relative to the number of sample institutions offering 4-year undergraduate and postgraduate programs. Second, the number of sample students in 4-year institutions was increased by 20 percent. Finally, the sample sizes of graduate students and professional students were slightly reduced. These three changes in the NPSAS sample design reflect the goal of following a large sample of bachelor's degree recipients through postgraduate experiences. Based on these changes, approximately 16,300 potential bachelor's degree recipients were identified for the first B&B cohort. These students were identified using institutionally provided lists of students who filed for graduation in the 1992–93 academic year.

All B&B-eligible sample members who completed the NPSAS interview were retained for future follow up. Of the 11,810 cases considered to be NPSAS completes, 11,254 were delivered with the first wave of data (designated as sample type 1). The remaining 556 were identified later as potentially eligible for B&B and were delivered as part of sample type 4. A subsample of approximately 10 percent of the remaining eligible cases with at least some data (either partial computer-assisted telephone interview (CATI) data, institution data, or parent data) was also identified and delivered as sample types 2 and 3. Additional NPSAS sample members (who were not part of the B&B cohort) were identified as potential bachelor's degree completers in the 1992–93 academic year based on review of the completed NPSAS institution information from the CATI nonrespondents.

All student NPSAS respondents (sample type 1) were included in the final B&B sample. The subsample selection was carried out by constructing a file of all B&B-eligible nonrespondents in sample types 2, 3, and 4. Complete cases, cases with pending interviewer appointments, sample members determined to be ineligible, and cases finalized as noninterviews were excluded from the subsampling file. This file was then sorted by institution stratum, student stratum, and student sample type in order to affect stratification in the selection process. A systematic sample of 200 persons was selected from approximately 450 in the file. At the start of interviewing, the final sample for the first B&B cohort numbered 12,478 recent graduates, consisting of: 11,254 NPSAS respondents classified as sample type 1; 300 student nonrespondents with NPSAS parent data (sample type 2); 164 other NPSAS nonrespondents (sample type 3); and 760 NPSAS respondents identified during the data processing phase as potentially eligible for B&B (sample type 4).

Transcripts for all sample members were requested from the NPSAS schools that awarded the bachelor's degrees. A total of 1,094 respondents who were either NPSAS noninterviews or who were otherwise deemed ineligible for B&B based on the telephone interview were reclassified as eligible based on transcript data.

After data collection for the first follow up was complete for both the interview and transcript components, additional cases in the initial sample were found to be ineligible for B&B. People were retained for follow up in later rounds if they were found to be eligible in either the CATI or the transcript component. Therefore, 10,080 CATI-eligible cases were retained for follow up plus an additional 1,094 transcript-eligible cases. In addition, 18 cases for which eligibility was unknown for both components were retained. All together, 11,192 cases were retained for future rounds.

Of these 11,192 B&B-eligible cases, 10,773 completed the 1992–93 NPSAS, 10,080 completed the first follow up (B&B:93/94), 10,976 had transcripts in B&B:93/94, 10,093 completed the second follow up (B&B:93/97). There were 9,274 cases which responded to all three CATI interviews through the second follow up.

Data Collection and Processing

B&B surveyed its first cohort—1992–93 bachelor's degree recipients—approximately 1 year after graduation, in 1994, and again in 1997. Both follow-up surveys were administered by the National Opinion Research Center (NORC) at the University of Chicago. The third follow up will be conducted in 2003 by Research Triangle Institute (RTI).

Reference dates. In the first follow up of the 1992–93 cohort, respondents were asked to provide their current enrollment status, employment status, and marital status as of April 1994. Similarly, respondents to the second follow up reported their status as of April 1997.

Data collection. Data are collected through student interviews and college transcripts. The data collection procedures for the follow ups of the first B&B cohort are described below.

Student interview. The first follow-up student interview was administered between June and December 1994. Sample members were initially mailed a letter containing information about the survey and a toll-free number they could call to schedule interviews. CATI began approximately 1 week later and was initiated in two waves. Wave 1 consisted of students who were respondents in the 1992–93 NPSAS or for whom parent data were available. Wave 2 consisted of students who were nonrespondents in the 1992–93 NPSAS and for whom no parent data were available. NPSAS respondents who were identified as potentially eligible for B&B during the NPSAS data processing phase were also included in Wave 2.

Telephone interviewing continued for a period of 16 weeks. All cases still pending after this time were sent to field interviewers to gather in-person information. A maximum of 14 calls was set, with a call defined as contact with the sample member, another person in the sample member's household, or an answering machine. After 14 calls, attempts to contact the sample member by telephone were terminated and the case was sent to field interviewers.

Methods of refusal conversion were tailored to address the reasons each member had given for nonparticipation, as determined by reviewing the call notes. Letters were sent to sample members addressing the specific reasons for their refusal (too busy, not interested, confidentiality issues, etc.). Following these mailings, a final phone interview was attempted from the central CATI site. Continuing refusals were forwarded to the field to be contacted in person by a field interviewer. The field staff was successful in completing 3,050 (82 percent) of these cases.

The second follow-up student interview was administered between April and December 1997. Sample members were initially mailed a letter and informational leaflet containing information about the survey, and a toll-free number and/or e-mail address through which they could obtain further information, schedule an interview, or provide an updated phone number. CATI began approximately 1 week later, and continued for 16 weeks. Cases pending at the end of this time were sent to field interviewers and worked from July through December 1997. Phone interviewers made 13, rather than 14, attempts to contact sample members. If phone interviewers had no success in the first 13 attempts, the case was forwarded to telephone case management specialists before being sent to field interviewers.

There were also slight modifications to the methods used to locate sample members. Prior to the beginning of the CATI, all cases had been sent to a credit bureau database service to obtain updated phone and address information about each sample member. Telephone numbers were also available from the previous interview (B&B:93/94 in 1997 or NPSAS in 1994) and the NCOA/Telematch update service NORC had used for all main survey respondent data in February, 1996, prior to the start of the field test. The "best" phone number was assumed to be the number most recently obtained.

Additional locating information used by locating specialists (in the order of their use) were: (1) all respondent-generated information (e-mails, address corrections from the U.S. Post Office, any previously acquired respondent phone numbers); (2) last known telephone number of the parent(s); (3) graduate schools (if applicable); (4) undergraduate institutions/alumna associations; (5) the other two credit bureau updating services; (6) military locating service if applicable; and (7) the Department of Motor Vehicles in the state which issued the respondent's last known driver's license.

A total of 1,679 respondents (15 percent of the total eligible sample) refused to complete the interview at some point in the process. After a 2-week "cooling off" period, these cases were contacted by trained interviewers experienced in refusal conversion. The CATI refusal converters were able to complete 335 of the refusal cases. Continuing refusals were forwarded to the field to be contacted in person by a field interviewer. A total of 3,993 cases (36 percent of the total sample) were sent to the field staff, which was successful in completing 2,954 (74 percent) of these cases.

Transcript component. In addition to data gathered from sample members, the B&B first follow up included a transcript component which attempted to capture student-level coursetaking and grades for eligible sample members. Transcripts were requested for all sample members from the NPSAS schools that awarded their bachelor's degrees.

Data collection for the first follow up began in August 1994, when transcript request packets were mailed to all 715 NPSAS sample schools from which B&B sample members graduated. In addition to student transcripts, schools were asked to provide a course catalog and information on their grading and credit-granting systems and their school term. A transcript was requested for all 12,478 students in the B&B sample, although not all transcripts were coded due to sample member ineligibility. Prompting of nonresponding schools began in September 1994 by the telephone center and attempts were made to address any concerns of school staff regarding confidentiality or the release of transcripts.

The design of the transcript processing system capitalized on work done in previous NORC studies. The process and flow system, however, was changed in four significant areas. First, since the sample of schools from which transcripts were collected was known, the system was designed around the school as the primary unit rather than around the student. Second, transcripts were entered after all school-level information about schedule, grading, and credit-granting systems was collected and verified. The system enforced these parameters and ensured that the transcripts were internally consistent within the school. Third, the transcript coders worked with the full transcript when entering and coding courses. This allowed the coders to view each entry in context and make intelligent, informed decisions when they encountered difficult situations. Finally, the system was designed so that course-level information within schools was entered only once; subsequent duplicate course entries were selected by the coder from a dynamic school-level list of all courses entered from previous transcripts. If a course failed to match a pre-existing entry, the coder searched the school-level table to see if other courses existed for the abbreviation. If a course was not in the table, the coder entered the full course title, the number of credits, and the grade.

Editing. Various edit checks, including CATI edits, have been used in processing B&B data; however, these have not been documented in B&B methodology reports.

Estimation Methods

Weighting is used in B&B to adjust for sampling and unit nonresponse. Imputation is used to estimate baseline weights from NPSAS when these data are missing. No imputation is performed on data collected in B&B follow ups. Procedures for the first B&B cohort are described below.

Weighting. Weights were modified from baseline weights in the 1992–93 NPSAS to adjust for nonresponse and the tighter eligibility criteria of the B&B sample. The 1992–93 NPSAS sample development and weights calculation documentation can be found in the *Sampling Design and Weighting Report for the 1993 National Postsecondary Student Aid Study.* (See section 7, Methodology and Evaluation Reports.)

After verifying sample eligibility against transcript data, sample members were stratified according to institutional type and student type. These strata reflected the categories used in the 1992–93 NPSAS, with some modifications. The 1992–93 NPSAS categorized schools into 22 institutional strata based on highest degree offered, control (public or private), for-profit status, and the number of degrees the institution awarded in the field of education (with schools subsequently designated "high ed" or "low ed"). For weighting purposes, these 22 institutional strata were collapsed in B&B to the 16 that granted baccalaureate degrees. The six NPSAS strata representing 2-year or less-than-2-year institutions were reclassified in B&B according to control and included within the correlative "4-year, bachelor's, low ed" stratum. This affected a total of 19 cases. The five student types originally identified in the 1992–93 NPSAS were collapsed to three in the B&B: baccalaureate business majors, baccalaureate other majors, and baccalaureate field unknown, resulting in 48 total cells.

Baseline weights for all B&B-eligible students were adjusted for final degree totals. Control totals for baccalaureate degrees awarded were calculated based on the Integrated Postsecondary Education Data Systems (IPEDS) Completions file for academic year 1992–93. The NPSAS institution sample frame was matched to the IPEDS file, and the total number of baccalaureate degrees awarded was calculated by institutional stratum. An adjusted weight was calculated for each case by multiplying the NPSAS base weight by the ratio of the sum of degrees awarded to the sum of the base weights for the appropriate institutional stratum. This weight became the B&B base weight.

In order to make nonresponse adjustments for weights, adjustment cells were created by cross-classifying cases by institutional stratum and student type. Each cell was checked to verify that it met two conditions: (1) the cell contained at least 15 students, and (2) the weighted response rate for the cell was at least two-thirds (67 percent) of the overall weighted response rate. Any cells that did not meet both conditions were combined into larger cells by combining two student type cells (baccalaureate business majors and "all other degrees") within the same institutional stratum. If this larger cell still did not meet the criteria specified above, all three student types from that institutional stratum were combined. Once all cells were defined, the B&B base weight variable (derived above) was multiplied by the inverse of the weighted response rate for the cell.

Final weights for the second follow up (B&B:93/97) were calculated, using a two-step process by making a nonresponse adjustment to the baseline B&B weight calculated for B&B:93/94. The 16 institutional-type and 3 student-type strata were used again, with the same process described previously.

Imputation. The sample for the first B&B cohort included 23 eligible cases for which the baseline weight from the 1992–93 NPSAS was equal to zero. Weights for these cases were imputed using the average of all nonzero baseline weights within the same institution at which the baccalaureate degree was attained. One of the cases with a missing weight happened to be the only representative of that institution. The baseline weight was imputed for this case by using the average across all nonzero weights within the same institutional stratum and student type cell.

There was no other imputation of data items in the three data collections of the first B&B cohort.

Future Plans

The next follow up of the first B&B cohort (1992–93 bachelor's degree recipients) will be conducted in 2003.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Taylor Series approximations are used to estimate standard errors in B&B.

Nonsampling Error

The majority of nonsampling errors in B&B can be attributed to nonresponse. Other sources of nonsampling error include: use of ambiguous definitions; differences in interpreting questions; inability or unwillingness to give correct information; mistakes in recording or coding data; and other instances of human error occurring during the multiple stages of a survey cycle.

Coverage error. The B&B sample is drawn from NPSAS. Consequently, any coverage error in the NPSAS sample will be reflected in the B&B. (Refer to chapter 16 for coverage issues in NPSAS.)

Nonresponse error. Overall response rates were very high for both follow ups of the 1992–93 B&B cohort. Data for unit and item nonresponse are broken down below.

Unit nonresponse. Of the 12,478 cases originally included in the first B&B sample, 1,520 were determined during the interview process to be ineligible or out of scope (primarily because their date of graduation fell outside the July 1–June 30 window). A total of 10,958 cases were considered to be eligible during the interviewing period of the B&B first follow up, and interviews were completed with 10,080 of these respondents, representing a 92 percent unweighted response rate.

Response rates were even higher for transcript collection. In all, 626 of 635 eligible schools complied with the request for transcripts, providing transcripts for 10,970 of the 12,478 cases—a 98 percent response rate.

In the second follow up, of the 11,192 cases identified as eligible B&B sample members, 30 were subsequently found to be out of scope or ineligible (29 were sample members who had died since 1993, and one case was identified as ineligible when it was determined the respondent had never received a baccalaureate degree). Interviews were completed with 10,970 of the 11,220 in-scope cases, for a final unweighted response rate of 90 percent. While response rates were similar across many demographic subgroups, some distinctive differences exist. Response rates decreased slightly with age (93.1 percent of those under 26 compared to 90.4 percent of those over 30 participated) but participation among males and females was approximately equal. Response rates were also similar among Whites, Blacks, and American Indians (ranging from 89.5 percent to 91.6 percent) but are substantially lower for Asians/Pacific Islanders (only 82.2 percent) and those identifying themselves as "other" (73.8 percent).

Table 8 summarizes the unit level and overall level weighted response rates across B&B administrations.

Item nonresponse. Of the more than 1,000 variables included in the final data set, 68 contain more than 10 percent missing data. The largest nonresponse was for items involving recollection of test scores and dates. Respondents also had difficulty recalling detailed information about undergraduate loans and loan payments when the respondent had more than three loans. The two primary sections of the survey, concerning postbaccalaureate education and employment, had very low rates of nonresponse.

Table 8. Unit level and overall level weighted response rates for selected B&B surveys

	Unit level weighted response rate			
	Base year 1st level	Base year 2nd level	1st wave	2nd wave
B&B – students	88.2	73.6	83.4	90.4
	Overall level weighted response rate			
	Base year 1st level	Base year 2nd level	1st wave	2nd wave
B&B – students	88.2	67.1	79.1	79.7

SOURCE: Seastrom, Salvucci, Walter, and Shelton (forthcoming). *A Review of the Use of Response Rates at NCES.*

Measurement error. Three sources of measurement error identified in B&B are respondent error, interviewer error, and error involved in the coding of course data from transfer schools where no school-level data were available.

Respondent error. Several weeks after the first follow-up interview of the 1992–93 cohort, a group of 100 respondents was contacted again for a reinterview. These respondents were asked a subset of items included in the initial interview to help assess the quality of those data. Results indicate that the questions elicited similar information in both interviews. Ninety-two percent of respondents gave consistent responses when asked if they had taken any courses for credit since graduating from college. Among the 8 percent with inconsistent responses, most had a short enrollment spell that they mentioned in the initial interview but not in the reinterview.

Ninety-six percent of respondents gave consistent information in both interviews when asked whether they had worked since graduation. Almost three-quarters of respondents gave the same number in both interviews when asked about the number of jobs they held since graduation; 26 percent gave inconsistent responses. Upon scrutiny, many of these discrepancies resulted from jobs held around the time of graduation that were reported in just one of the interviews. Although respondents were asked to include jobs that began before graduation if they ended after graduation, confusion over whether to include such jobs accounted for many of the inconsistencies noted in the reinterview. The 1993–94 B&B field test also included a reinterview study. (See *Measurement Error Studies at the National Center for Education Statistics*, NCES 97–464.)

Interviewer error. The monitoring procedure for statistical quality control used in B&B extends the traditional monitoring criteria (which focus specifically on interviewer performance) to an evaluation of the data collection process in its entirety. This improved monitoring system randomly selects active work stations and segments of time to be monitored, determines what behaviors will be monitored and precisely how they will be coded, and allows for real-time performance audits, thereby improving the timeliness and applicability of corrective feedback and enhancing data quality. Results for the first follow up of the 1992–93 B&B cohort revealed a low rate of interviewer error, about three errors for every 100 minutes monitored.

Quality control procedures are also established for field interviewing. The first two interviewer-administered completed questionnaires are sent to a field manager for editing. These cases are edited and logged, and appropriate feedback is given to the interviewer. Additionally, 10 percent of these cases whether administered over the phone or in person are validated by field managers. When deemed necessary, the field managers continue to edit additional cases to monitor data quality. The need for additional monitoring is based on the field manager's subjective judgment of the field interviewer's skill level. As with the edited cases, validated cases are logged and reported weekly.

Transfer school course coding. The first follow up of the 1992–93 B&B cohort included a transcript data collection. Although transcripts were requested only from the institution awarding the baccalaureate degree, transcripts from previous transfer schools were often attached. Course data from these transfer school transcripts were coded, but no attempt was made to collect additional information from these schools. Due to the lack of school-level information on the 1,938 transfer schools involved, data from these transcripts are not the same quality as data coded from the baccalaureate institution's transcripts.

Data Comparability

At present, data are only available for the B&B first and second follow-up surveys conducted in 1994 and 1997. There are no current comparable data available.

6. CONTACT INFORMATION

For content information on B&B, contact:

Aurora M. D'Amico
Phone: (202) 502–7334
E-mail: aurora.d'amico@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Baccalaureate and Beyond Longitudinal Study: 1993/94 First Follow-up Methodology Report, NCES 96–149, by P.J. Green, S.L. Meyers, P. Giese, J. Law, H.M. Speizer, and V.S. Tardino. Washington, DC: 1996.

Baccalaureate and Beyond Longitudinal Study: 1993/97 Second Follow-up Methodology Report, NCES 1999–159, by P. Green, S. Myers, C. Veldman, and S. Pedlow. Washington, DC: 1999.

Survey Design

Baccalaureate and Beyond Longitudinal Study: Second Follow-up Field Test Report, 1996, NCES 97–261, by C. Veldman, P.J. Green, S. Myers, L. Chuchro, and P. Giese. Washington, DC: 1997.

Baccalaureate and Beyond First Follow-up Field Test Report, 1993 (B&B:93/94), NCES 94–371, by P.J. Green, H.M. Speiger, and B.K. Campbell. Washington, DC: 1994.

Sampling Design and Weighting Report for the 1993 National Postsecondary Student Aid Study, by R.W. Whitmore, M.A. Traccarella, and V.G. Iannacchione. Research Triangle Park, NC: Research Triangle Institute, 1995.

Data Quality and Comparability

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, and M. Saba. Washington, DC: 1997.

A Review of the Use of Response Rates at NCES (forthcoming), by M. Seastrom, S. Salvucci, E. Walter, and K. Shelton.

Chapter 19: Survey of Earned Doctorates (SED)

1. OVERVIEW

The Survey of Earned Doctorates (SED) is an annual census of new doctorate recipients from accredited colleges and universities in the United States. SED is funded by five federal agencies: the National Science Foundation (lead sponsor), the Department of Education, the Department of Agriculture, the National Institutes of Health, and the National Endowment for the Humanities.

Only research doctorates—primarily Ph.D.s, Ed.D.s, and D.Sc.s—are counted in SED. Professional doctorates (e.g., M.D., J.D., Psy.D.) are excluded. While the graduate schools are responsible for distributing SED forms to students, the surveys are completed by the doctorate recipients themselves. Collected information includes demographic characteristics of recipients, educational history from high school to doctorate, sources of graduate school support, debt level, and postgraduation plans.

The first SED was conducted during the 1957–58 academic year. In addition to housing the results of all surveys, the Doctorate Records File (DRF)—the survey database—contains public information on earlier doctorate recipients back to 1920. Thus, the DRF is a virtually complete data bank on more than 1.3 million doctorate recipients. The DRF also serves as the sampling frame for the biennial Survey of Doctorate Recipients (SDR), a longitudinal survey of science, engineering, and humanities doctorates employed in the United States.

ANNUAL CENSUS OF NEW RESEARCH DOCTORATE RECIPIENTS

SED collects self-reported data on:
- Demographic characteristics
- Educational history from high school to doctorate
- Mechanisms of financial support in graduate school
- Debt related to education
- Postgraduation plans

Purpose
To obtain consistent, annual data on individuals receiving research doctorates from U.S. institutions for the purpose of assessing trends in Ph.D. production.

Components
There is one component to SED.

Survey of Earned Doctorates. The doctorate institution is responsible for distributing the surveys to research doctoral candidates and collecting the surveys for mailback to the contractor. The doctorate recipients themselves complete the surveys. The following information is collected in SED: all postsecondary institutions attended and years of attendance; all postsecondary degrees received and years awarded (although only the *first* baccalaureate, master's, first-professional, and doctorate degrees are entered in the database); years spent as a full-time student in graduate school; specialty field of doctorate; type of financial support during graduate school; level of debt incurred in undergraduate and graduate school; employment/study status in the year preceding doctoral award; postgraduation plans (how definite, study vs. employment, location); high school location and year of graduation; demographic characteristics (sex, race/ethnicity, date and place of birth, citizenship status, country of citizenship for non-U.S.

citizens, marital status, number of dependents, disability status, educational attainment of parents); and personal identifiers (name, Social Security Number, and permanent address). The following information is keyed as verbatim text but only coded upon special request: dissertation title, dissertation field, and department (or interdisciplinary committee, center, etc.) that supervised the doctoral program.

Periodicity

Annual since inception of SED in the 1957–58 academic year. The database also includes basic information (obtained from public sources) on doctorates for the years 1920 to 1957.

2. USES OF DATA

The results from SED are used by government agencies, academic institutions, and industry to address a variety of policy, education, and human resource issues. The survey is invaluable for assessing trends in doctorate production and the characteristics of Ph.D. recipients. SED data are used to monitor the educational attainment of women and minorities, particularly in science and engineering. The increasing numbers of foreign citizens earning doctorates in the United States are studied by country of origin, field of concentration, sources of graduate school support, and U.S. "stay" rate after graduation. Trends in time-to-doctorate are also analyzed by field, type of support received, and personal characteristics such as marital status. The data on postdoctoral plans provide insight into the labor market for new Ph.D.s, and the careers of new Ph.D.s can be followed in the longitudinal Survey of Doctorate Recipients, whose sample is drawn from SED.

There is also substantial interest in the institutions attended by Ph.D.s. Doctorate-granting institutions frequently compare their survey results with peer institutions, and undergraduate institutions want to know their contribution to doctorate production. The availability of Carnegie Classifications in the DRF facilitates meaningful comparisons of the institutions attended by the different demographic groups (e.g., men vs. women). Separate indicators for historically Black colleges and universities can allow researchers to examine the roles these play in the educational attainment of Blacks.

3. KEY CONCEPTS

Some of the key terms and analytic variables in SED are described below.

Research Doctorate. Any doctoral degree that (1) requires the completion of a dissertation or equivalent project of original work (e.g., musical composition), and (2) is not exclusively intended as a degree for the practice of a profession. While the most typical research doctorate is the Ph.D., there are more than 50 other degree types (e.g., Ed.D., D.Sc., D.P.A., D.B.A.). Not included in this definition are professional doctorates: M.D., D.D.S., D.V.M., O.D., D.Pharm., Psy.D., J.D., and other similar degrees.

Doctorate-granting Institution. Any postsecondary institution in the United States that awards research doctorates (as defined above) and that is accredited at the higher education level by an agency recognized by the Secretary of the U.S. Department of Education. There are about 400 doctorate-granting institutions.

Field of Doctorate. Specialty field of doctoral degree, as reported by the doctorate recipient. There are about 280 fields on the SED Specialties List, grouped under the following umbrellas: agricultural sciences; biological sciences; health sciences; engineering; computer and information sciences; mathematics; physical sciences (subdivided into astronomy, atmospheric science and meteorology, chemistry, geological and related sciences, physics, and miscellaneous physical sciences); psychology; social sciences; humanities (subdivided into history, letters, foreign languages and literature, and other humanities); education; and professional fields (subdivided into business management and administrative services, communications, and other professional fields). Because field of doctorate is designated by the doctorate recipient, the classification in SED may differ from that reported by the institution in the NCES IPEDS Completions Survey. (See chapter 14.)

Time-to-doctorate. There are two standard, published measures of time-to-doctorate. *Total time-to-degree (TTD)* measures the total elapsed time between baccalaureate and doctorate, including time not enrolled in school. TTD can only be computed if baccalaureate year is known. *Registered time-to-degree (RTD)* gauges the time in attendance at all colleges and universities between receipt of the baccalaureate and doctoral award, including years of attendance not related to the doctoral program. RTD can only be computed if all years of attendance after the baccalaureate have been provided. Both of these measures are computed from several items in the educational history section of the questionnaire.

Source of Support. Any source of financial support received during graduate school. Doctorate recipients are asked to mark all types of support received and also to indicate the primary and secondary sources of support. For most SED years, sources are categorized as own/family resources; university-related (teaching and research assistantships, university fellowships, college work-study); federal research assistantships (by agency); other federal support (by mechanism and agency); nonfederal U.S. nationally competitive fellowships (by funding organization); student loans (Stafford, Perkins); and other sources (business/employer, foreign government, state government).

In 1997–98, the number of source options was reduced from 35 to 13. Sources are no longer identified by the specific provider (e.g., federal agency, foundation, type of loan) since students do not always have that knowledge. Only the mechanism of support (e.g., fellowship, research assistantship, loan) is now requested. Most current categories are aggregates of multiple categories on previous questionnaires (e.g., the new category "research assistantship" (RA) combines five earlier categories—university-related RA, NIH RA, NSF RA, USDA RA, and other federal RA). The following three categories are new as of 1997–98: dissertation grant, internship or residency, and personal savings.

4. SURVEY DESIGN

Target Population

All individuals awarded research doctorates from accredited colleges and universities in the United States between July 1 of one year and June 30 of the following year. There are currently about 43,000 research doctorates awarded annually by nearly 400 institutions located in the 50 states and Puerto Rico. Institutions in other U.S. territories do not grant research doctorates.

Sample Design

SED is a census of all recipients of research doctorates.

Data Collection and Processing

The data collection and editing process spans an 18-month period ending 6 months after the last possible graduation date (i.e., June 30). The update of the database and preparation of tables for first data release generally require another 4–6 months. From inception of SED in 1957–58 through the 1995–96 cycle, the survey was conducted by the National Research Council (NRC) of the National Academy of Sciences. The 1996–97 SED was collected by the NRC and processed by the new contractor, the National Opinion Research Center (NORC) of Chicago. NORC will conduct future administrations through the 2000–01 SED. The 1996–97 and 1997–98 administrations are considered a ***transition period.*** Not all NRC procedures were implemented during this period, and NORC continues to develop and test new procedures.

Reference dates. The data are collected for an academic year, which includes all graduations from July 1 of one year through June 30 of the following year.

Data collection. In advance of each survey, the contractor staff reviews the listings of accredited U.S. institutions in the *Higher Education Directory* to confirm that past participants are still doctorate-granting and identify accredited institutions that are newly doctorate-granting. As further confirmation of doctorate-granting status, the degree levels offered are checked on the IPEDS Institutional Characteristics (IC) File. (See chapter 14.) By July of each year, questionnaires are mailed to the institutions for distribution to doctoral candidates who expect to receive their degree between July 1 and June 30 of the following year. Institutional Coordinators are responsible for the distribution, collection, and return of the surveys. They are asked to provide official graduation lists or commencement programs along with the questionnaires, and to provide addresses for students who did not complete questionnaires.

Upon receipt of a graduation batch, the contractor staff compares the names of students on completed questionnaires ("self reports") with the names in the commencement program or on the official graduation list. Any discrepancies are followed up with the institution for confirmation of graduation. If an address for a nonrespondent is provided by the institution or found through other means, a letter and questionnaire are mailed to the individual to request completion of the survey. A second attempt is made to elicit participation if a response is not received within a month. In recent years, these efforts have yielded enough completed surveys to increase the survey's overall self-report rate by 5–7 percentage points.

For doctorate recipients still missing survey returns after these mailings, "skeleton" records are created from information contained in commencement programs or on graduation lists: name; doctorate institution, field, and year; similar information for baccalaureate and master's degrees; and sex (if it can be positively assumed from the name). Skeleton records have accounted for 4.1 to 8.2 percent of the records each year during the 1990s. In addition, a small percentage of surveys every year (usually less than 1 percent) are classified as "institutional" returns, having been completed by the institutions with whatever information was available to them. While institutional returns may contain more information than is available from commencement programs, the information is minimal compared to the self-reported surveys.

Staff undergo intensive training in the complexities of coding and checking procedures, and are monitored throughout the collection cycle.

Data processing. SED processing includes two special efforts to increase response rates for key items. The data entry procedures used by both the NRC and NORC include triggers if any of eight "critical" items is missing: date of birth, sex, citizenship status, country of citizenship (if foreign), race/ethnicity, baccalaureate institution, baccalaureate year, and postdoctoral location. If any of these items is absent, a "missing information letter" (MIL) is generated and sent to the respondent. For these cases, five noncritical items (if missing) are also requested: birth place, high school graduation year, high school location, master's institution, and year of master's degree.

A second follow-up effort requests the same critical items from the doctorate-granting institutions, both for individuals who never completed a survey (skeletons) and for individuals who completed a survey (self-reports) but did not return the MIL. Because of the lower MIL yield during the transition period, more information was requested from institutions in 1996–97 and 1997–98.

Editing. Records are processed through a multilayered edit routine that checks all variables for valid ranges of values and reviews the interrelationships among variables. The NRC performed these edits and the correction of errors online during data entry; then the full data file was processed a second time through selected edits after survey closure. NORC's CADE system also includes built-in range edits, but the interrelationship (consistency) edits are done after CADE is completed and after derived variables are created. There are more than 200 edit tests for SED: about 85 range edits (all hard, mandatory edits that cannot be overridden), and nearly 120 interrelationship edits. About two-thirds of the interrelationship edits are hard edits. The remaining third are soft edits, which can be overridden after the responses are double-checked and verified as accurate.

The entire battery of edit tests was reviewed during the 1994–95 SED cycle. A large set of interrelationship tests was developed at this time to verify the accuracy of foreign-country coding for the various time frames covered in the survey. Other interrelationship tests check for reasonable time frames in the doctorate recipient's chronology, from date of birth through date of doctoral award. Still others verify that the appropriate items are answered in a skip pattern (e.g., study vs. employment postdoctoral plans).

Estimation Methods

No weighting is performed since SED is a census. Some logical assumptions are made during coding and updating of the database. For example, U.S. citizenship is assumed for Ph.D.s who designate their ethnicity as Puerto Rican since, legally, Puerto Ricans are U.S. citizens. Entries of "China" in country of citizenship may be recoded to either Taiwan or the People's Republic of China, based on the locations of birth place, high school, baccalaureate institution, and master's institution. Postdoctoral plans are assumed to be employment if items in the employment section are answered and the postdoctoral study section is blank. Postdoctoral study is assumed if the opposite scenario is indicated.

Recent Changes

During the 1990s, the National Science Foundation asked NRC to implement several new procedures in an effort to improve both the quantity and quality of SED data. Beginning with the 1989–90 SED, there has been rigorous follow-up of complete nonrespondents and respondents who did not answer key data items. Race/ethnicity, postdoctoral location, and country of citizenship (if foreign) were first followed up in the 1989–90 cycle, increasing the completeness of these items from that time forward. In the mid-1990s, more than 100 new edit tests were implemented to check the coding of certain foreign countries for specific time frames. In the 1995–96 cycle, the survey instrument was reformatted to make it more respondent-friendly; although content remained the same, the survey form was expanded from 4 to 12 pages.

During the 1996–97 cycle, the contract for conducting SED was transferred from the NRC to NORC; this has brought some changes in procedures, as documented in earlier sections. In addition, the 1997–98 questionnaire included a major revision to the source of support question; the response set has been changed from specific providers *and* mechanisms of support to only mechanisms. The marital status question was also changed in

1997–98 to (1) separate "widowed" from "separated/divorced" and (2) add a new category for "living in a marriage-like relationship."

Future Plans

Additional changes to SED are under consideration, both to capture new data relevant to current issues in graduate education and to collect better data through existing questions.

5. DATA QUALITY AND COMPARABILITY

The 1990s brought a reexamination of all operational processes, introduction of state-of-the-art technologies, evaluations of data completeness and accuracy, and renewed efforts to attain even higher response rates for every item in the survey. A Technical Advisory Committee was established to guide the conduct of SED with a look toward the future. A Validation Study was conducted to assess the limitations of SED data, and data user groups were convened to advise on survey content. The survey instrument was reformatted to make it more respondent-friendly, and questions are now being revised to collect more complete and accurate information. While the transition from one contractor to another has caused some reduction in the completeness of the data, efforts are underway to return response rates to their earlier levels and to further enhance the quality of the available data.

Sampling Error

SED is a census and, thus, is not subject to sampling error.

Nonsampling Error

The main source of nonsampling error in SED is measurement error. Coverage error is believed to be very limited. Unit and item response rates have been very high and relatively stable since the first survey in 1957–58 (although somewhat lower during the transfer of SED administration to the new contractor).

Coverage error. SED is administered to a universe of research doctorates identified by the universe of research doctorate-granting institutions. Therefore, undercoverage might result from (1) an incomplete institution universe, and/or (2) an incomplete enumeration of research doctorates. SED coverage has been evaluated and found to be less than 1 percent, due to the high visibility of doctorate-granting institutions and a comprehensive approach to data collection.

Every year, the universe of institutions is reviewed and compared to the institutional listings in the *Higher Education Directory* and other sources to determine the current list of doctorate-granting institutions. Any institutions newly determined to be doctorate-granting are contacted for verification of doctorate-granting status and then invited to participate in SED. A few qualifying institutions refuse to participate, but it is known from the IPEDS Completions Survey that these institutions contribute minimally to the overall doctorate population.

Individual doctorate recipients are enumerated through (1) survey forms completed by the new Ph.D.s and returned by the institution; (2) transmittal rosters that provide the official count of doctorates, the number of surveys completed and returned, and the names of individuals who did not complete surveys; and (3) commencement programs covering every graduation at an institution over the course of a year. Comparisons of the number of research doctorates in SED with the total number of doctorates reported by institutions in NCES' IPEDS Completions Survey show that SED's coverage differs by less than 1 percent.

Nonresponse error. Targets have been set for both unit and item response in SED. While the target rates are not always attained, response has been unusually high for a mail survey throughout the 40+ years of SED.

Unit nonresponse. Basic information on nonrespondents can be obtained from institutions or commencement programs, so records exist for all recipients of research doctorates. However, response to SED is measured by the percentage of doctorate recipients who complete the surveys themselves (*self-report rate*), thus providing details that are not available from any other source. SED's goal is a stable self-report rate of 94–95 percent. This rate has been achieved or surpassed in all but 14 of the 41 surveys processed to date (through the 1997–98 SED). Response first fell below the target rate in 1986 and stayed low throughout the rest of the 1980s, at which time site visits and intensive follow-up procedures were initiated in an effort to increase the percentage of self-reported questionnaires. Response achieved the target level from 1990 to 1995 but has since fallen below target (92.8 percent in 1996 and about 91.5 percent in 1997 and 1998).

Because SED is administered through the doctorate-granting institutions, the self-report rate is dependent upon their overall cooperation and survey practices. In the 1997–98 SED, nearly one-third (31 percent) of the 387 institutions had self-report rates below 90 percent, which is the target rate for institutions. Nonresponse tends to be concentrated in a small group of institutions. In the 1997–98 SED, the 10 institutions with the largest numbers of doctorate nonrespondents (ranging from 51 to 131) accounted for 40.4 percent of the total self-report nonresponse that year.

To improve tracking of institution response rates, NORC has devised an "early warning system" to identify institutions whose self-report rates lag behind the goal of 90 percent. Estimates for each seasonal graduation are developed, based on the numbers for an institution's graduations in previous years. This system also allows monitoring of institutions with specific substantive interest for SED (e.g., engineering schools, institutions awarding doctorates to large numbers of racial/ethnic minorities).

Item nonresponse. Certain items are available for all doctorate recipients, whether or not they completed a questionnaire: name, doctorate institution, field of doctorate, month and year of doctoral award, and type of doctorate. This information is always provided by the institution in its commencement program or graduation list.

A 95 percent target is set for eight "critical" items: date of birth, sex, citizenship, country of citizenship (if foreign), race/ethnicity, baccalaureate institution, baccalaureate year, and postdoctoral location. From the 1989–90 SED (when rigorous follow up of these items began) to the 1995–96 SED, all items but postdoctoral location achieved response rates above 95 percent. Rates for all critical items except sex and foreign country of citizenship fell below goal in the 1996–97 and 1997–98 SED administrations, the transition period between contractors. Decreases in item response during this period ranged from 2.5 percentage points for race/ethnicity to 4.8 points for baccalaureate year. These decreases stemmed in part from parallel decreases in the overall self-report rates for these two survey cycles and in part from less intensive follow-up efforts during the transition period. However, the higher level of valid data in the 1997–98 SED, as compared to the previous year, suggests a return to increased item response.

"Critical" items are followed up through letters to self-reporting survey respondents and through requests to institutions for Ph.D.s who did not complete questionnaires. Thus, the response rates for these items often exceed the overall self-report rate for the survey. Because information can be obtained from sources other than the doctorate recipients, item response rates for SED are computed on the universe of recipients, whether or not they responded to the survey.

The target rate for all "noncritical" survey items is 90 percent. During much of the past decade, most noncritical items achieved goal or were within 2 percentage points. Fewer items attained a 90 percent response during the recent transition period between contractors. The results for the 1997–98 SED showed 27 of the 49 noncritical items achieving the 90 percent target and 22 items with response rates below target. Throughout SED's history, a few items have had, and will continue to have, lower response rates because they are not applicable to all individuals (e.g., master's degree information, secondary work activity). Other items with lower-than-average response rates relate to timelines from college entrance to doctoral graduation, the most complex segment of the questionnaire.

Some items with below-goal response in the first half of the 1990s surpassed the 90 percent target once the questionnaire was reformatted for the 1995–96 SED. The 1995–96 survey form was expanded from 4 to 12 pages, allowing instructions to be clarified and multipart questions to be broken out into separate, more distinguishable questions.

Although the questionnaire reformat has been successful in many areas, declines in response to key demographic items (citizenship, foreign country of citizenship, and race/ethnicity) and Social Security Number (the critical linking variable) are of concern. Decreases in response rates were relatively small in the 1995–96 SED, but response subsequently dropped to the levels of the 1980s during the transition from one contractor to another. As of the 1995–96 SED, the demographic items are asked at the end of the survey; these items were located at the beginning of the survey in all earlier years.

Measurement error. Most measurement error in SED results from respondents' misinterpretation of questions or limited recall of past events. The 1994 Validation Study sought to determine the limitations of SED data. Think-aloud interviews were conducted with recent Ph.D. recipients, who were asked to complete a second survey form within a few months of their original survey submission. The question on sources of support caused the most difficulty; few Ph.D.s responded exactly the same as in the initial survey. Problems with this item were confirmed by focus group discussions and comparisons of SED results with raw data obtained from organizations that fund the various types of support. The source of support question was revised in the 1997–98 SED to request only the mechanism of support (e.g., research assistantship, fellowship, loan) rather than the actual source of funding (e.g., NSF, NIH), which some students do not know.

Interviewees were sometimes confused about the educational history section of the survey, thinking that short-term attendance at a school or attendance not leading to a degree was not required. Others were unsure about whether or not to include the time spent working on the dissertation. Such inconsistencies have an impact on time-to-doctorate computations. To address these issues, several new questions on time to degree were added to the 2001 SED.

Several interviewees also had difficulty responding to the questions on postgraduation plans because, although they currently had a job, they wanted to indicate that they were still seeking a position that would satisfy their aspirations. These comments led to discussions among sponsors and other data users about the intent of the postdoctoral questions and what information is most relevant for policymaking.

Data Comparability

Because a prime use of SED data is trend analysis, tremendous efforts have been made to maintain continuity of survey content. Only three new items have been added since 1973: disability status, number of years as a graduate student, and debt level at time of doctorate receipt. However, occasional changes have been made to the response categories for an item, sometimes affecting the comparability of the data over time. For the items on disability status and debt level, such changes occurred frequently enough to make comparisons for the early years unreliable.

The second modification to the 1997–98 questionnaire affects the sources of support item. The response set was overhauled to request information on only the mechanism of support (e.g., research assistantship, fellowship, loan) rather than mechanism *and* funder (e.g., NIH RA, NSF RA, university fellowship, NSF fellowship, Ford Foundation fellowship, Stafford loan, Perkins loan). As noted under Measurement Error above, focus groups and interviews revealed that students do not always know the actual source of their support, particularly when the funder is the federal government. The 1997–98 response set for the item on sources of support also includes three new categories: dissertation grant, internship/residency, and personal savings.

This major change has broken the time series for the sources of support item except for selected sources. NORC mapped the pre-1998 response categories to the new response set and then compared the 1997–98 distribution of responses to earlier distributions back to 1990. Significant shifts were observed in the proportions for some categories—raising concerns about whether the new code frame accurately captures the desired information on sources of support (e.g., tuition remission), and also suggesting the need for more cognitive work in this area. Therefore, ***users should be cautious about making generalizations regarding the financing of doctoral education over time.***

Another comparability issue for SED involves changes (generally additions) over the years to the survey's Specialties List, which is used to code fields for degrees, postdoctoral study, and employment. Because any specialties added to the list would have been coded into an "other" category (e.g., other biological sciences) in previous surveys, users should be careful in their interpretation of time-series field data at the most disaggregated level. The historical changes in the Specialties List are documented in *Science and Engineering Doctorates: 1960–91* (NSF 93–301), and the subsequent series, *Science and Engineering Doctorate Awards* (NSF 00–304).

While both unit and item response rates in SED have been relatively stable through the years, fluctuations can affect data comparability. This is especially important to consider when analyzing data by citizenship and race/ethnicity, where very small fluctuations in response may result in increases or decreases in counts that do not reflect real trends. New procedures implemented in the early 1990s had a significant positive impact on response to these two items, as well as to the items on foreign country of citizenship and postdoctoral location, making the data from 1990 to 1996 better in both quantity and quality than data from the late 1980s. Item response for citizenship and race/ethnicity have fallen to the level of 1990 and earlier years, and item response for postdoctoral location is lower than most years in the 1990s. However, while response to country of citizenship among non-U.S. citizens fell 3 percentage points in the first transition year (the 1996–97 SED), it returned to pretransition levels in the 1997–98 SED.

The reformat of the questionnaire in 1995–96, described in earlier sections, resulted in substantial increases in response to primary source of support, postdoctoral work activity, and postdoctoral employment field. Users should take these changes into account when analyzing trends.

Comparisons with IPEDS. The IPEDS Completions Survey also collects data on doctoral degrees, but the information is provided by institutions rather than by doctorate recipients. The number of doctorates reported in the IPEDS Completions Survey is slightly higher than in SED. This difference is largely attributable to the inclusion in the IPEDS Completions Survey of nonresearch doctorates, primarily in the fields of theology and education. The differences in counts have been generally consistent since 1960, with ratios of IPEDS-to-SED counts ranging from 1.01 to 1.06. Because a respondent to SED may not classify his/her specialty identically to the way the institution reports the field in the IPEDS Completions Survey, differences between the two surveys in the number of doctorates for a given field may be greater than the difference for all fields combined.

6. CONTACT INFORMATION

The National Science Foundation is the Systems Manager of Record for the Survey of Earned Doctorates. The micro-data can be used by institutions that enter into Licensing Agreements with NSF. The persons to contact concerning this are:

Susan Hill, Director
Doctorate Data Project
National Science Foundation
(703) 292-7790

Ron Fecso, Chief Statistician
Division of Science Resources Statistics
National Science Foundation
(703) 292-7769

For content information about SED, contact:

NCES/USED Contact:

Nancy Borkow
Phone: (202) 502-7311
E-mail: nancy.borkow@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006-5651

NSF Contact:

Susan T. Hill
Phone: (703) 292-7790
E-mail: sthill@nsf.gov

Mailing Address:

Human Resources Statistics Program
Division of Science Resources Statistics, Room 965 S
National Science Foundation
4201 Wilson Boulevard
Arlington, VA 22230

NORC Contact:

Lance Selfa
Phone: (312) 759-4031
E-mail: selfa@norcmail.uchicago.edu

Mailing Address:

Doctorate Records Project
National Opinion Research Center (NORC)
55 East Monroe Street
Chicago, IL 60603

7. METHODOLOGY AND EVALUATION REPORTS

General

National Science Foundation. *Guide to NSF Science and Engineering Resources Data*, NSF 95-318, by Carolyn F. Shettle. Arlington, VA: 1995. [Updated information can be found at http://www.nsf.gov/sbe/srs/ssed/sedmeth.htm.]

Survey Design

National Opinion Research Center. *Report on Cognitive Research for the 2000 SED Questionnaire Development Task*, by B. Dugoni, L. Lee, and A. Baldwin. Chicago: 1999.

Policy Research Methods, Inc. *Report on Cognitive Research for the 2000 SED Questionnaire Development Task*. Arlington, VA: 1996.

Data Quality and Comparability

National Opinion Research Center. *Evaluation Report 1998: Quality Profile for the 1997–1998 Survey of Earned Doctorates*. Chicago: 1999.

National Research Council. *Evaluation Report 1996: Quality Profile for the 1995–1996 Survey of Earned Doctorates*. Washington, DC: 1997.

National Research Council. *Validation Study of the Survey of Earned Doctorates*, by L. Ingram and P. Ries. Washington, DC: 1994.

Chapter 20: National Assessment of Educational Progress (NAEP)

1. OVERVIEW

The National Assessment of Educational Progress (NAEP) is mandated by Congress to assess the educational accomplishments of U.S. students and monitor changes in those accomplishments. As the only nationally representative and continuing assessment of what America's students know and can do in selected subject areas, NAEP serves as the "Nation's Report Card." The *main* NAEP regularly assesses the achievements of students in grades 4, 8, and 12 at the national level. The *state* NAEP assessed at both grades 4 and 8 in at least one subject in 1992, 1996, 1998, 2000, 2002, and 2003. In 2003 and beyond, State NAEP is planning to assess in at least two subjects, reading and mathematics, every 2 years at grades 4 and 8. The *trend* NAEP tracks national long-term trends in science, mathematics, and reading at ages 9, 13, and 17. It tracked writing proficiency trends at grades 4, 8, and 11 through 1999, when critical issues were identified with having so few writing prompts. The national assessments were first implemented in 1969 and were conducted on an annual or biennial basis through 1995, and annually since 1996. The state assessments have been administered biennially since 1990.

In 1988, Congress established the National Assessment Governing Board (NAGB) to provide policy guidance for the execution of NAEP. NAGB is composed of national and local elected officials, chief state school officers, classroom teachers, local school board members, leaders of the business community, and others. Specifically, it is charged by Congress to select subject areas to be assessed; identify appropriate achievement goals for each age group; develop assessment objectives; design the methodology of the assessment; and produce guidelines and standards for national, regional, and state comparisons.

BIENNIAL SURVEY OF A SAMPLE OF ELEMENTARY/ SECONDARY STUDENTS

Three assessments:
- Main National NAEP
- Main State NAEP
- Trend NAEP

Four component surveys:
- Elementary and Secondary School Students Survey
- School Characteristics and Policies Survey
- Teacher Survey
- SD/LEP Survey / Excluded Student Survey

Purpose

To (1) monitor continuously the knowledge, skills, and performance of the nation's children and youth; and (2) provide objective data about student performance at national, regional, and, since 1990, state levels.

Components

NAEP comprises three separate assessments: *main national, main state,* and *trend.* Each of these assessments consists of four components: Elementary and Secondary School Students Survey; School Characteristics and Policies Survey; Teacher Survey; and Students with Disabilities or Limited English Proficiency (SD/LEP) Survey (for the *main* NAEP) or Excluded Student Survey (for the *trend* NAEP). In 1985, the Young Adult Literacy Study was also conducted nationally as part of NAEP, under a grant to the Educational Testing Service and Response Analysis Corporation; this study assessed the literacy skills of 21- to 25-year-olds. In addition, a High School Transcript Study is periodically conducted as a component of NAEP. (See chapter 28.)

In 1996, 1998, and 2000, the national main and state assessments included a special study of the effects of accommodations on the performance of students with special needs. A subsample of students with disabilities or limited English proficiency was given special accommodations (e.g., extended testing time) during the assessment. A comparison subsample took the assessment under standard conditions. Both subsamples met the 1996 criteria for inclusion of special needs students in NAEP.

National-level Assessments. The *main national* NAEP and *trend* NAEP are both designed to report information for the nation and specific geographic regions of the country (Northeast, Southeast, Central, and West). However, these two assessments use separate samples of students from public and nonpublic schools: grade samples for the *main national* NAEP (4th, 8th, 12th grades), and age/grade samples for the *trend* NAEP (age 9/grade 4; age 13/grade 8; age 17/grade 11). The test instruments for the two assessments are based on different frameworks, student and teacher background questionnaires vary, and the results for the two assessments are reported separately. (See *Elementary and Secondary School Students Survey* below for the subject areas assessed.)

The assessments in the *main* NAEP follow the curriculum frameworks developed by NAGB and use the latest advances in assessment methodology. The test instruments are flexible so they can be adapted to changes in curricular and educational approaches. Recent assessment instruments for the main NAEP have been kept stable for short periods of time, allowing short-term trends to be reported from 1990 through 2003.

To reliably measure change over longer periods of time, the *trend* NAEP must be used. For long-term trends, past procedures must be precisely replicated with each new assessment, and the survey instruments do not evolve with changes in curricula or educational practices. The instruments used today for the trend NAEP are identical to those developed in the mid-1980s. The trend NAEP allows measurement of trends from 1969 to the present.

State-level Assessments. The *main state* NAEP was implemented in 1990 on a trial basis and has been conducted biennially since that time. (The assessments from 1990 to 1994 are referred to as trial state assessments, or TSAs.) Participation of the states was completely voluntary until 2001. The reauthorization of the Elementary and Secondary Education Act, also referred to as the "No Child Left Behind" legislation, requires states that receive Title I funding to participate in state NAEP assessments in reading and mathematics at grades 4 and 8 every 2 years. State participation in other state NAEP subjects (i.e., science and writing) remains voluntary. Separate representative samples of students are selected for each jurisdiction to provide that jurisdiction with reliable state-level data concerning the achievement of its students. The state assessment included nonpublic schools only in 1994, 1996, and 1998. This practice ended because of low participation rates. See below for the subject areas assessed.

Elementary and Secondary School Students Survey. The primary data collected by NAEP relate to student performance and educational experience as reported by students. Major assessment areas include: reading, writing, mathematics, science, civics, U.S. history, geography, social studies, and the arts.

In 1988, the *main national* NAEP assessed student performance in reading, writing, civics, and U.S. history, and conducted small special-interest assessments in geography and document literacy. In 1990, it assessed mathematics, writing, and science; in 1992, reading, mathematics, and writing; in 1994, reading, U.S. history, and world geography; and in 1996, science and mathematics. A probe of student performance in the arts at grade 8 was conducted in 1997. Reading, writing, and civics were assessed in 1998. (*Trend* NAEP was assessed in 1999.) In 2000, the main national NAEP assessed mathematics and science and, for 4th graders only, reading. In 2001, history and geography were assessed, and in 2002, reading and writing. In 2003, the assessments are in reading and mathematics for 4th and 8th graders.

The subjects assessed in *trend* NAEP are mathematics, science, reading, and until 1999, writing. The biennial assessments from 1988 through 1996 covered all subjects. The next trend assessment will be conducted in 2004 and then trend assessments are scheduled to be administered every 4 years.

Representative *main state-level* data were collected for the first time in the 1990 trial state assessment, when 8th-grade students were assessed in mathematics. In 1992, state-level data were collected in 4th-grade reading and mathematics, and in 8th-grade mathematics. In 1994, 4th-grade reading was assessed. In 1996, 4th-grade mathematics and 8th-grade mathematics and science were assessed. The 1998 NAEP collected state-level data in reading at grades 4 and 8, and writing at grade 8. The 2000 NAEP assessments covered mathematics and science, the 2002 assessments covered reading and writing, and the 2003 assessments cover reading and mathematics.

The student survey also asks questions about the student's background, as well as questions related to the subject area and the student's motivation in completing the assessment. *Student background questions* gather information about race/ethnicity, school attendance, academic expectations, and factors believed to influence academic performance, such as homework habits, the language spoken in the home, and the quantity of reading materials in the home. Some of these questions document changes that occur over time, and remain unchanged over assessment years.

Student subject-area questions gather three categories of information: time spent studying the subject, instructional experiences in the subject, and perceptions about the subject. Because these questions are specific to each subject area, they can probe in some detail the use of specialized resources such as calculators in mathematics classes.

Students are also asked how often they have been asked to write long answers to questions on tests or assignments that involved (this subject). In earlier assessments, students were also asked how many questions they thought they answered correctly, how difficult they found the assessment, how hard they tried on this test compared to how hard they had tried on most other tests or assignments they had taken that year in school, and how important it was to them to do well on this test. (In 2003, NAEP dropped the motivation questions.)

School Characteristics and Policies Survey. This survey collects supplemental data about school characteristics and school policies that can be used analytically to provide context for student performance issues. School data include: enrollment, absenteeism, dropout rates, curricula, testing practices, length of school day and year, school administrative practices, school conditions and facilities, size and composition of teaching staff, tracking policies, schoolwide programs and problems, availability of resources, policies for parental involvement, special services, and community services.

Teacher Survey. This survey collects supplemental data from teachers whose students are respondents to the assessment surveys. Part I of the Teacher Questionnaire covers background and general training, requesting information on the teacher's race/ethnicity, sex, age, years of teaching experience, certification, degrees, major and minor fields of study, coursework in education, coursework in specific subject areas, amount of in-service training, extent of control over instructional issues, and availability of resources for the classroom. Part II of the Teacher Questionnaire covers training in the subject area and classroom instructional practices, specifically the teacher's exposure to issues related to the subject and the teaching of the subject, pre- and in-service training, ability level of the students in the class, length of homework assignments, use of particular resources, and how students are assigned to particular classes.

SD/LEP Survey. This survey is completed in the *main* NAEP assessments by teachers of students selected to participate in NAEP but classified as having disabilities (SD) or classified as limited English proficient (LEP). Information is collected on the background and characteristics of each SD/LEP student and the reason for the SD/LEP classification, as well as whether these students receive accommodations in district or statewide tests. For SD students, questions ask about the student's functional grade levels and special education programs. For LEP students, questions ask about the student's native language, time spent in special language programs, and the level of English language proficiency. This survey is used to determine whether the student should take the NAEP assessment. If any doubt exists about a student's ability to participate in the assessment, the student is included. Beginning with the 1996 assessments, NAEP has allowed accommodations for both SD and LEP students.

Excluded Student Survey. This survey is completed in the *trend* NAEP for students who are sampled for the assessment but excluded by the school. Following exclusion criteria used in previous trend assessments, a school can exclude students with limited English-speaking ability, students who are educable mentally retarded, and students who are functionally disabled—if the school judges that these students are unable to "participate meaningfully" in the assessment. This survey is only completed for those students who are actually excluded from the assessment (whereas the SD/LEP Survey in the main assessment is also completed for participating students who are SD or LEP students—see above).

High School Transcript Study. Transcript studies have been conducted in 1987, 1990, 1994, 1998, and 2000. The studies collect information on current course offerings and course-taking patterns in the nation's schools. Transcript data can be used to show course-taking patterns across years that may be associated with proficiency in subjects assessed by NAEP. Transcripts are collected from grade 12 students in selected schools from the NAEP sample. (For more information, see chapter 28, Other NCES Surveys and Studies.)

Special Studies. The 1998 assessment included three subsamples that used special procedures to study specific aspects of writing and civics. The special studies samples were drawn from the grade-only population. The three special studies consisted of: (1) Writing – 50: a sample of students in grades 8 and 12 who received 50-minute writing blocks in assessments sessions where no other writing format was administered; (2) Writing – Classroom: a sample of students in grades 4 and 8 who were assessed based on written assignments the students had completed as part of their regular school curriculum; and (3) Civics – Special Trend: a sample of students in grades 4, 8, and 12 who were assessed using the booklets and testing conditions used in the 1988 civics assessment.

Oral Reading Study Assessment. In 2002, NAEP conducted a special study on oral reading. The NAEP 2002 Oral Reading Study looked at how well the nation's 4th graders can read aloud a grade-appropriate story. NAEP assessed a random sample of 4th-grade students selected for the NAEP 2002 reading and writing assessments. The assessment provided information about a student's fluency in reading aloud and examined the relationship between oral reading accuracy, rate (or speed), fluency, and reading comprehension.

Technology-Based Assessment (TBA) Project. TBA was designed with five components—three empirical studies (Mathematics Online, Writing Online, and Problem Solving in Technology-Rich Environment), a conceptual paper (Computerized Adaptive Testing), and an online school and teacher questionnaire segment, which is already operational. The primary goals of Mathematics Online (MOL) are to understand how computer delivery affects the measurement of NAEP math skills, to gain insights into the operational and logistical mechanics of computer-delivered assessments, and to evaluate the ability of 4th and 8th graders to deal with mathematics assessments delivered on computer. At grade 8, an additional goal is to investigate the technical feasibility of generating alternate versions of multiple-choice and constructed-response items using an "on-the-fly" (OTF) technology. MOL was field tested in 2002. The Writing Online (WOL) study is intended to help NAEP learn how computer delivery affects the measurement of NAEP performance-based writing skills, to gain insights into the operational and logistical mechanics of computer-delivered writing assessments, and to evaluate the ability of 8th graders to deal with writing assessments delivered on computer. WOL was field tested in 2002. The Problem Solving in Technology-Rich Environment (TRE) study was designed to develop an example set of modules to assess problem solving using technology. These example modules will use the computer to present multimedia tasks that cannot be delivered through conventional paper-and-pencil assessments, but which tap important emerging skills. TRE is being field tested in 2003.

Periodicity

Annual from 1969 to 1979 and, again, beginning in 1996; biennial in even-numbered years from 1980 to 1998. A probe of 8th graders in the arts area was conducted in 1997. State-level assessments, first initiated in 1990, follow the same schedule as the national assessments. Prior to 1990, NAEP was required to assess reading, mathematics, and writing at least once every 5 years. The previous legislation required assessments in reading and mathematics at least every 2 years, in science and writing at least every 4 years, and in history or geography and other subjects selected by the National Assessment Governing Board at least every 6 years. The No Child Left Behind Act requires NAEP to conduct national and state assessments at least once every 2 years in reading and mathematics in grades 4 and 8. In addition, in the future, NAEP will conduct a national assessment and may conduct a state assessment in reading and mathematics in grade 12 every 4 years starting in 2005. Finally, to the extent that time and money allow, NAEP will be conducted in grades 4, 8, and 12 at regularly scheduled intervals in additional subjects including writing, science, history, geography, civics, economics, foreign languages, and arts.

2. USES OF DATA

NAEP serves as the Nation's Report Card. It is the only ongoing, comparable, and representative assessment of what American students know and can do in several subject areas. Policymakers are keenly interested in NAEP results because they address national outcomes of education, specifically the level of educational achievement. In addition, state-level data, available for many states since 1990, allow both state-to-state comparisons and comparisons of individual states with the nation as a whole.

During NAEP's history, more than 200 reports across 12 subject areas have provided a wealth of information on students' academic performance, learning strategies, and classroom experiences. Together with the performance results, the basic descriptive information collected about students, teachers, administrators, and communities can be used to address the following educational policy issues:

- Instructional practices: What instructional methods are being used?
- Students-at-risk: How many students appear to be at-risk in terms of achievement, and what are their characteristics? What gaps exist between at-risk categories of students and others?
- Teacher workforce: What are the characteristics of teachers of various subjects?
- Education reform: What policy changes are being made by our nation's schools?

However, ***users should be cautious in their interpretation of NAEP results. While NAEP scales make it possible to examine relationships between students' performance and various background factors, the relationship that exists between achievement and another variable does not reveal its underlying cause, which may be influenced by a number of other variables.*** NAEP results are most useful when they are considered in combination with other knowledge about the student population and the educational system, such as trends in instruction, changes in the school-age population, and societal demands and expectations.

NAEP materials such as frameworks and released questions also have many uses in the educational community. Frameworks present and explain what experts in a particular subject area consider important. Several states have used NAEP frameworks to revise their curricula. After most assessments, NCES releases nearly one-third of the questions to the interested public. Released constructed-response questions and their corresponding scoring guides have served as models of innovative assessment practices in the classroom.

3. KEY CONCEPTS

The achievement levels for NAEP assessments are defined below. For subject-specific definitions of achievement levels and additional terms, refer to NAEP *Technical Reports*, *Report Card* reports, and other publications.

Achievement Levels. Starting with the 1990 NAEP, the NAGB developed achievement levels for each subject at each grade level to measure how well students' *actual* achievement matches the achievement *desired* of them. The three levels are:

Basic. Partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.

Proficient. Solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.

Advanced. Superior performance. This level is only attained by a very small percentage of students (3–6 percent) at any of the three grade levels assessed.

4. SURVEY DESIGN

Target Population

Students enrolled in public and nonpublic schools in the 50 states and the District of Columbia, who are deemed assessable by their school and classified in defined grade/age groups—grades 4, 8, and 12 for the *main national* assessments, and ages 9, 13, and 17 for the *trend* assessments in science, mathematics, and reading. Grades 4 and/or 8 are usually assessed in the *state* NAEP; the number of grades has varied in the past, depending on availability of funding (although testing for 4th and 8th graders in reading and mathematics every 2 years is now required for states that receive Title I funds). Only public schools were included in the *state* NAEP prior to 1994 and after 1998.

Sample Design

The sample for each NAEP assessment is selected using a complex multistage clustered design involving the sampling of students from selected schools within selected geographic areas, called primary sampling units (PSUs), across the United States. The sample designs for NAEP assessments have been similar since the mid-1980s. In 1983, student samples were expanded to include both age- and grade-representative populations. Since 1988, the samples have been drawn from the universe of 4th, 8th, and 12th graders for the Elementary and Secondary School Students Survey; from the teachers of those students for the Teacher Survey; and from the school administrators at those elementary and secondary schools for the School Characteristics and Policies Survey. In 1996, SD/LEP students were oversampled for a special study of SD/LEP inclusion; hence, exclusion rules and availability of accommodations were different than in previous studies. The national-level sample for each NAEP assessment contains approximately 7,000 to 10,000 students for each grade assessed—or 0.42 percent of the national student population for each grade.

NAEP's multistage sampling process involves the following steps:

(1) Selection of PSUs

(2) Selection of schools (public and nonpublic) within the selected PSUs

(3) Assignment of session types to schools

(4) Selection of students for session types within the selected schools

In 1996, the special study of SD/LEP inclusion required an additional step for the main assessments: the assignment of "sample types" to schools based on specific criteria for excluding students with limited English proficiency or severe disability, and the provision or nonprovision of accommodations. Results from this study indicated that revising the criteria for including students had little impact on the numbers of students included. Because of the lack of impact, the revised criteria for including students will be used in future assessments. Provision of accommodations was found to have a limited impact on performance results. NAEP made a full transition to providing allowable accommodations to all students who need them in 2002.

Selection of PSUs. In the first stage of sampling, the United States (the 50 states and the District of Columbia) is divided into geographic PSUs. The PSUs are classified into four regions (Northeast, Southeast, Central, and West), each containing about one-fourth of the U.S. population. In each region, PSUs are additionally classified as metropolitan or nonmetropolitan, resulting in eight subuniverses of PSUs.

For the 1998 main assessment, 94 PSUs were selected; 22 of these PSUs were designated as certainty units because of their size. Within each major stratum (subuniverse), further stratification was achieved by ordering the noncertainty PSUs according to several additional socioeconomic characteristics (e.g., median household income, educational level of residents over 25 years of age, demographic characteristics). One PSU was selected from each of the 72 noncertainty strata, with probability proportional to size (total population from the 1990 census). To enlarge the samples of Black and Hispanic students, thereby enhancing the reliability of estimates for these groups, PSUs from the high-minority strata were sampled at twice the rate of PSUs from the other strata. This was achieved by creating smaller strata with high-minority subuniverses.

There were no long-term trend NAEP samples in 1998; however, in 1996, when 94 PSUs were selected for the main assessment, 52 PSUs were selected for the long-term trend samples. Of these 52 trend PSUs, 10 selected with certainty because of their size, 6 were selected from the 12 remaining main sample certainty PSUs, and 36 were selected from the 72 noncertainty strata independently of the main sample selection.

Selection of schools. In the second stage of sampling, public schools (including Bureau of Indian Affairs—BIA—schools and Department of Defense Education Activity—DODEA—schools) and nonpublic schools (including Catholic schools) within each of the selected PSUs are listed according to the grades associated with the three age classes: age class 9 refers to age 9 or grade 4 in the trend NAEP or grade 4 in the main NAEP; age class 13 refers to age 13 or grade 8 in the trend NAEP or grade 8 in the main NAEP; age class 17 refers to age 17 or grade 11 in the trend NAEP or grade 12 in the main NAEP.

The school lists are obtained from two sources. Regular public, BIA, and DODEA schools are obtained from the school list maintained by Quality Education Data, Inc. (QED). Catholic and other nonpublic schools are obtained from the NCES Private School Survey. (See chapter 3.) To ensure that the state samples provide an accurate representation, public schools are stratified by urbanization, minority enrollment, and median household income. Nonpublic schools are stratified by type of control (e.g., parochial, private), urban status, and enrollment per grade. Once the stratification is completed, the schools within each PSU are assigned a probability of selection that is proportional to the number of students per grade in each school.

An independent sample of schools is selected separately for each age/grade so that some schools are selected for assessment of two age/grades and a few are selected for all three. Schools within each PSU are selected (without replacement) with probabilities proportional to assigned measures of size. Nonpublic schools and schools with high minority enrollment are oversampled.

The manner of sampling schools for the long-term trend assessments is very similar to that used for the main assessments. The primary difference is that nonpublic schools and schools with high minority enrollment are not oversampled. Schools are not selected for both main and long-term trend assessments at the same age/grade.

Assigning sample type to schools. As noted earlier, schools in the 1996 ***main*** assessments were assigned a "sample type" based on specific criteria for excluding students, with the goal of determining the effect of different exclusion criteria in NAEP assessments. Historically, a small proportion (less than 10 percent) of the sampled students have been excluded from NAEP assessments because they are SD/LEP students whom their local schools determined could not take the assessments. In recent years, increased attention has been given to including as many of these students as possible in NAEP assessments.

Three different sample types were assigned to the schools selected for the 1996 main assessment. For sample type 1 schools, the exclusion criteria for the main samples were identical to those used in 1990 and 1992. Sample type 2 schools used new inclusion criteria for SD and LEP students. In sample type 3 schools, the new inclusion criteria were used and, in addition, accommodations were offered to SD and LEP students. The specific criteria and availability of accommodations varied among the schools. The most frequently provided accommodations were small group administration, extended time (untimed testing), and, in mathematics, bilingual assessment booklets. Sample type was assigned separately for each grade.

In the 1998 national main and state reading assessments, sample types 2 and 3 were assigned to schools. The writing and civics assessments were administered to sample type 3 schools only.

Assignment of session types to schools. In the third stage of sampling, assessment sessions are assigned to the selected schools found to be in-scope, with three aims in mind. The first is to distribute students to the different session types (e.g., assessment in a particular academic subject or pilot test of new items) across the whole sample for each age class so that the target numbers of assessed students will be achieved. The second is to maximize the number of different session types that are administered within a given selected school without violating minimum session sizes. The third is to give each student an equal chance of being selected for a given session type regardless of the number of sessions conducted in the school. Beginning in 2002, for the main assessment, session types were no longer assigned to schools; rather, sessions all had a common session design so that multiple subjects can be spiraled across students.

Selection of students. The fourth stage of sampling involves random selection of national samples representing the entire population of U.S. students in grades 4, 8, and 12 for the main assessment and the entire population of students at ages 9, 13, and 17 for the long-term trend assessment (grades 4, 8, and 11 for the writing assessment). The selection process differs slightly based on whether the sample of students is needed for the main national assessment, the long-term trend assessment, or the main state assessment. A small number of students selected for participation are excluded because of limited English proficiency or severe disability.

To facilitate the sampling of students, a consolidated list is prepared for each school of all grade-eligible and age-eligible students (long-term trend assessments) or all grade-eligible students (main assessments) for the age class for which the school is selected. A systematic selection of eligible students is made from this list—unless all students are to be assessed—to provide the target sample size.

For example, to oversample Black and Hispanic students from public schools with low minority enrollment, as was done in 1998, after the initial sample was selected, the nonselected Black and Hispanic students were identified and listed. If the number of nonselected students was less than the number of selected students, then all nonselected Black and Hispanic students were assessed. Otherwise, Black and Hispanic students were sampled so that their overall within-school probability of selection was twice the rate of other students. Likewise in 1998, in each school where oversampling of SD/LEP students was to occur, the initial desired sample of students was drawn for each session assigned from the full list of eligible students. Among those students not selected for either of the two prior sampling operations for that school, the SD/LEP students were identified. A sample from among these was drawn, using a sampling rate that would achieve the double sampling rate required overall.

For schools assigned more than a single session type, which is the vast majority of schools, students are assigned to one of the various session types using specified procedures.

For each age class (separately for long-term trend and main samples), maxima are established as to the number of students who are to be selected for a given school. In those schools that, according to information on the sampling frame, have fewer eligible students than the established maxima, each eligible student enrolled at the school is selected in the sample for one of the sessions

BEST COPY AVAILABLE

assigned to the school. In other schools, a sample of students is drawn and students are assigned to sessions as appropriate. No student is assigned to more than one session. The maximum sample sizes are established in terms of the number of grade-eligible students (by sample type in 1996) for the main samples, and in terms of the number of students in each age class for the trend samples.

The classroom-based writing study involved the random selection of one English/language arts classroom from each 4th- and 8th-grade school in which a writing assessment was to be conducted. At the same time, the students in that classroom were listed on a writing study linkage form so that the classroom students who also took the national writing assessment could be identified. The classroom's English/language arts teacher was asked to work with the students and have them select two examples of their best classroom writing. The students were asked to answer a few questions about each selection. The teachers completed an interview with the supervisor who collected the writing materials after the assessment.

Excluded students. Some students are excluded from the student sample because they are deemed unassessable by school authorities. The exclusion criteria for the main samples differ somewhat from those used for the long-term trend samples. In order to identify students that should be excluded from the *main* assessments, school staff members are asked to identify those SD or LEP students who do not meet the NAEP inclusion criteria. School personnel are asked to complete an SD/LEP questionnaire for all SD and LEP students selected into the NAEP sample, whether they participate in the assessment or not. For the long-term *trend* assessments, excluded students are identified for each age class, and an Excluded Student Survey is completed for each excluded student.

For the special study of SD/LEP inclusion in the 1996 main assessment, oversampling procedures were applied to SD/LEP students at all three grades in sample types 2 and 3 for mathematics and in sample type 3 for science.

Main national and state NAEP sample sizes. Not all subject areas are assessed in every assessment year. In 1998, the main national NAEP assessed students in reading, writing, and civics at all three grades. The main state NAEP in 1998 assessed students in writing at grade 8 and in reading at grades 4 and 8. The total target sample size for the 1998 state assessments was 396,000 (132,000 for each grade and subject). The sample included students from an average of 225 schools per state. For the main national NAEP, the total target sample size was 132,000 students from 2,000 schools nationwide. Sample sizes by grade ranged from 8,000 to 13,000 in reading; from 20,000 to 26,000 in writing; and from 6,000 to 8,000 in civics. A separate civics trend sample included 2,000 students from each grade.

In comparison, the 1996 main national assessment, which tested mathematics and science at all three grade levels, required fewer than 100,000 students from about 1,800 schools. The state-level assessment, which tested only two grade levels, required a total sample of about 350,000 students from nearly 10,000 schools because of the number of states that participated.

Long-term trend NAEP sample sizes. The long-term trend assessment tested the same four subjects across years through 1999, using relatively small national samples. Samples of students were selected by age (9, 13, and 17) for mathematics, science, and reading, and by grade (4, 8, and 11) for writing. Students within schools were randomly assigned to either mathematics/science or reading/writing assessment sessions subsequent to their selection for participation in the assessments. The next long-term trend assessment will be administered in 2004, and then every 4 years thereafter (but not in the same years as the main assessments) in reading and mathematics.

Assessment Design

Since 1988, the NAGB has selected the subjects for the main NAEP assessments. NAGB also oversees creation of the frameworks that underlie the assessments and the specifications that guide the development of the assessment instruments.

Development of framework and questions. NAGB uses an organizing framework for each subject to specify the content that will be assessed. This framework is the blueprint that guides the development of the assessment instrument. The framework for each subject area is determined through a consensus process involving teachers, curriculum specialists, subject-matter specialists, school administrators, parents, and members of the general public.

Unlike earlier multiple-choice instruments, current instruments dedicate a majority of testing time to constructed-response questions that require students to compose written answers. Constructed-response questions provide a separate means of assessing ability, tapping recall not recognition.

The questions and tasks in an assessment are based on the subject-specific frameworks. They are developed by teachers, subject-matter specialists, and testing experts, under the direction of NCES and its contractors. For each subject-area assessment, a national committee of experts provides guidance and reviews the questions to ensure that they meet the framework specifications. For each state-level assessment, state curriculum and testing directors review the questions that will be included in the NAEP state component.

Matrix sampling. Several hundred questions are typically needed to reliably test the many specifications of the complex frameworks that guide NAEP assessments. However, administering the entire collection of cognitive questions to each student would be far too time consuming to be practical. Matrix sampling allows the assessment of an entire subject area within a reasonable amount of testing time (e.g., 50 minutes to an hour and a half). By this method, different portions from the entire pool of cognitive questions are printed in separate booklets and administered to different but equivalent samples of students. About 2,600 students respond to each block of items.

The type of matrix sampling used by NAEP is called focused, balanced incomplete block (BIB) spiraling. The NAEP BIB design varies according to subject area.

Data Collection and Processing

Since 1983, NCES has conducted NAEP through a series of contracts, grants, and cooperative agreements with the Educational Testing Service (ETS) and other contractors. ETS is directly responsible for developing the assessment instruments, analyzing the data, and reporting the results. Westat selects the school and student samples, trains assessment administrators, and manages field operations (including assessment administration and data collection activities). NCS Pearson is responsible for printing and distributing the assessment materials and for scanning and scoring students' responses.

Reference dates. Data for the main national NAEP and main state NAEP are collected at overlapping times during winter. Data for the long-term trend NAEP are collected during fall for age 13/grade 8; during winter of the same school year for age 9/grade 4; and during spring for age 17/grade 11.

Data collection. Until 2002, NCES relied heavily on school administrators for the conduct of main state NAEP assessments. Beginning with the 2002 assessments, NAEP contract staff conduct all NAEP assessment sessions. Obtaining the cooperation of the selected schools requires substantial time and energy, involving a series of mailings that includes letters to the chief state school officers and district superintendents to notify the sampled schools of their selection; additional mailings of informational materials; and introductory in-person meetings where procedures are explained.

The questionnaires for the School Characteristics and Policies Survey, the Teacher Survey, and the SD/LEP Survey are sent to the schools ahead of the assessment date so that they can be collected when the assessment is administered. Questionnaires not ready at this time are retrieved later, either through a return visit by NAEP personnel or through the mail.

NCS Pearson produces the materials needed for NAEP assessments. NCS Pearson prints identifying bar codes and numbers for the booklets and questionnaires, preassigns the booklets to testing sessions, and prints the booklet numbers on the administration schedule. These activities improve the accuracy of data collection and assist with the spiraled distribution process.

Assessment exercises are administered either to individuals or to small groups of students by specially trained field personnel. For all three ages in the long-term trend NAEP, the science and mathematics questions were administered using a paced audiotape. Beginning in 2004, the long-term trend assessments will be administered through test booklets read by the students.

For the long-term trend assessments, Westat hires and trains approximately 85 field staff to collect the data. Starting with the 2002 main national and state assessments, Westat has employed and trained about 3,000 field staff to carry out the assessments.

Westat ensures quality control across states by monitoring 25 percent of the sessions. Security of assessment materials and uniformity of administration are high priorities. (To date, there have been no reports from quality control monitors of serious breaches in procedures or major problems that could jeopardize the validity of the assessment.) After each session, Westat staff interview the assessment administrators to receive their comments and recommendations. As a final quality control step, a debriefing meeting is held with the state supervisors to receive feedback that will help improve procedures, documentation, and training for future assessments.

Data processing. NCS Pearson handles all receipt control, data preparation and processing, scanning, and scoring activities for NAEP. Using an optical scanning machine, NCS Pearson staff scan the multiple-choice selections, the handwritten student responses, and other data provided by students, teachers, and administrators.

An intelligent data entry system is used for resolution of the scanned data, the entry of documents rejected by the scanning machine, and the entry of information from the questionnaires. An image-based scoring system introduced in 1994 virtually eliminates paper handling during the scoring process. This system also permits online monitoring of scoring reliability and creation of recalibration sets.

ETS and NCS Pearson develop focused, explicit scoring guides with defined criteria that match the criteria emphasized in the assessment frameworks. The scoring guides are reviewed by subject area and measurement specialists, the Instrument Development Committees, NCES, and NAGB to ensure consistency with both question wording and assessment framework criteria. Training materials for scorers include examples of student responses from the actual assessment for each performance level specified in the guides. These exemplars help scorers interpret the scoring guides consistently, thereby ensuring the accurate and reliable scoring of diverse responses.

The image scoring system allows scorers to assess and score student responses online. This is accomplished by first scanning the student response booklets, digitizing the constructed responses, and storing the images for presentation on a large computer monitor. The range of possible scores for an item also appears on the display; scorers click on the appropriate button for quick and accurate scoring. The image scoring system facilitates the training and scoring process by electronically distributing responses to the appropriate scorers and by allowing ETS and NCS Pearson staff to monitor scorer activities consistently, identify problems as they occur, and implement solutions expeditiously. The system also allows the creation of calibration sets that can be used to prevent drift in the scores assigned to questions. This is especially useful when scoring large numbers of responses to a question (e.g., more than 30,000 responses per question in the main state NAEP). In addition, the image scoring system allows all responses to a particular exercise to be scored continuously until the item is finished, thereby improving the validity and reliability of scorer judgments.

The reliability of scoring is monitored during the coding process through (1) backreading, where table leaders review about 10 percent of each scorer's work to confirm a consistent application of scoring criteria across a large number of responses and across time; (2) daily calibration exercises to reinforce the scoring criteria after breaks of more than 15 minutes; and (3) a second scoring of 25 percent of the items appearing only in the main national assessment and 6 percent of the items appearing in both the main national and state assessments, and a comparison of the two scores to give a measure of interscorer reliability. To monitor agreement across years, a random sample of 20–25 percent of responses from previous assessments (for identical items) is systematically interspersed among current responses for rescoring. If necessary, current assessment results are adjusted to account for any differences.

To test scoring reliability, constructed-response item score statistics are calculated for the portion of responses that are scored twice. Cohen's Kappa is the reliability estimate used for dichotomized items and the intraclass correlation coefficient is used as the index of reliability for nondichotomized items. Scores are also constructed for items that are rescored in a later assessment. For example, some reading, writing, and civics items from 1994 were rescored in 1998. See the table below.

Editing. The first phase of data editing takes place during the keying or scanning of the survey instruments. Machine edits verify that each sheet of each document is present and that each field has an appropriate value. The edit program checks each booklet number against the

Table 9. Sample score ranges and percent agreements for constructed-response reading items that were scored twice

	Dichotomously scored items		Polytomously scored items	
	Cohen Kappa	Percent agreement	Intraclass correlation	Percent agreement
1998 national main assessment reading items				
4th grade	0.705–0.970	87–98	0.821–0.957	78–91
8th grade	0.665–0.996	84–100	0.761–0.977	64–98
12th grade	0.596–0.967	83–100	0.668–0.992	66–97
1994 reading items rescored in 1998				
4th grade	0.722 to 0.944	86–96	0.855 to 0.968	78–92
8th grade	0.678 to 0.983	83–99	0.798 to 0.978	64–96
12th grade	0.535 to 0.952	76–98	0.698 to 0.974	62–95

SOURCE: Derived from tables in appendix C, Allen, Donoghue, and Schoeps, *The NAEP 1998 Technical Report* (NCES 2001-509).

session code for appropriate session type, the school code against the control system record, and other data fields on the booklet cover for valid ranges of values. It then checks each block of the document for validity, proceeding through the items within the block. Each piece of input data is checked to verify that it is of an acceptable type, that the value falls within a specified range of values, and that it is consistent with other data values. At the end of this process, a paper edit listing of data errors is generated for nonimage and key-entered documents. Image-scanned items requiring correction are displayed at an online editing terminal.

In the second phase of data editing, experienced editing staff review the data errors detected in the first phase of editing, compare the processed data with the original source document, and indicate whether the error is correctable or noncorrectable per the editing specifications. Suspect errors found to be correct as stated but outside the edit specifications are passed through modified edit programs. For nonimage and key-entered documents, corrections are made later via key-entry. For image-processed documents, suspect errors are edited online. The edit criteria for each item in question appear on the screen along with the suspect item, and corrections are made immediately. Two different people view the same suspect data and operate on it separately, and a "verifier" ensures that the two responses are the same before the system accepts that item as correct.

For assessment items that must be paper-scored rather than scored on the image system (as was the case for some mathematics items in the 1996 NAEP), the score sheets are scanned on a paper-based scanning system and then edited against tables to ensure that all responses were scored with one and only one valid score, and that only raters qualified to score an item were allowed to score it. Any discrepancies are flagged and resolved before the data from that scoring sheet are accepted into the scoring system.

In addition, a count-verification phase systematically compares booklet IDs with those listed in the NAEP Administration Schedule to ensure that all booklets expected to be processed were actually processed. Once all corrections are entered and verified, the corrected records are pulled into a mainframe data set and then re-edited with all other records. The editing process is repeated until all data are correct.

Estimation Methods

Once NAEP data are scored and compiled, the responses are weighted according to the sample design and population structure and then adjusted for nonresponse. This ensures that the students' representation in NAEP matches their actual proportion of the school population in the grades assessed. The analyses of NAEP data for most subjects are conducted in two phases: scaling and estimation. During the scaling phase, item response theory (IRT) procedures are used to estimate the measurement characteristics of each assessment question. During the estimation phase, the results of the scaling are used to produce estimates of student achievement (proficiency) in the various subject areas. The marginal maximum likelihood methodology is then used to estimate characteristics of the proficiency distributions. Estimates of cognitive ability are included in the NAEP database. Estimates of other variables are not included in the database.

Weighting. The weighting for the national and state samples reflects the probability of selection for each student in the sample, adjusted for school and student nonresponse. The weight assigned to a student's responses is the inverse of the probability that the student would be selected for the sample. Through poststratification, the weighting ensures that the representation of certain subpopulations correspond to figures from the U.S. Census and the Current Population Survey (CPS).

Student base weights. The base weight assigned to a student is the reciprocal of the probability that the student was selected for a particular assessment. This probability is the product of the following four factors:

- the probability that the PSU was selected;
- the conditional probability that the school was selected, given the PSU;
- the conditional probability, given the selected samples of schools in the PSU, that the school was allocated the specified assessment; and
- the conditional probability, given the school, that the student was selected for the assessment.

Nonresponse adjustments of base weights. The base weight for a selected student is adjusted by two nonresponse factors. The first factor adjusts for sessions that were not conducted. This factor is computed separately within classes formed by the first three digits of PSU strata. Occasionally, additional collapsing of classes is necessary to improve the stability of the adjustment factors, especially for the smaller assessment components.

The second factor adjusts for students who failed to appear in the scheduled session or makeup session. This nonresponse adjustment is completed separately for each assessment. For assessed students in the trend samples, the adjustment is made separately for classes of students based on subuniverse and modal grade status. For assessed students in the main samples, the adjustment classes are based on subuniverse, modal grade status, and race class. In some cases, nonresponse classes are collapsed into one to improve the stability of the adjustment factors.

Scaling. For purposes of summarizing item responses, ETS developed a scaling technique that has its roots in Item Response Theory (IRT) and the theories of imputation of missing data.

The first step in scaling is to determine the percentage of students who give various responses to each cognitive, or subject-matter, question and each background question. For cognitive questions, a distinction is made between missing responses at the end of a block (i.e., missing responses subsequent to the last question the student answered) and missing responses prior to the last observed response. Missing responses before the last observed response are considered intentional omissions. Missing responses at the end of the block are generally considered "not reached" and treated as if the questions had not been presented to the student. In calculating response percentages for each question, only students classified as having been presented that question are used in the analysis. Each cognitive question is also examined for differential item functioning (DIF). DIF analyses identify questions on which the scores of different subgroups of students at the same ability level differ significantly.

Development of scales. Separate subscales are derived for each subject area. For the main assessments, the frameworks for the different subject areas dictate the number of subscales required. In the 1996 NAEP, five subscales were created for the main assessment in mathematics (one for each mathematics content strand), and three subscales were created for science (one for each field of science: earth, physical, and life). A composite scale is also created as an overall measure of students' performance in the subject area being assessed (e.g., mathematics). The composite scale is a weighted average of the separate subscales for the defined subfields or content strands. For the long-term trend assessments, a separate scale is used for summarizing proficiencies at each age/grade level in each of the subject areas—science, mathematics, reading, and writing.

Within-grade vs. cross-grade scaling. Reading and mathematics main NAEP assessments were developed with a cross-grade framework, where the trait being measured was conceptualized as cumulative across the grades of the assessment. Accordingly, a single 0-to-500 scale was established for all three grades in each assessment. In 1993, NAGB determined that future NAEP assessments should be developed using within-grade frameworks and be scaled accordingly. This both removes the constraint that the trait being measured is cumulative and eliminates the need for overlap of questions across grades. Any questions that happen to be the same across grades are scaled separately for each grade, thus making it possible for common questions to function differently in the separate grades.

The 1994 history and geography assessments were developed and scaled within-grade, according to NAGB's new policy. The scales were aligned so that grade 8 had a higher mean than grade 4, and grade 12 had a higher mean than grade 8. The 1994 reading assessment, however, retained a cross-grade framework and scaling. All three main assessments in 1994 used scales ranging from 0 to 500.

The 1996 long-term trend assessments converted to within-grade, using a 0 to 500 scale. The 1996 main science assessment was also developed within-grade, but adopted new scales ranging from 0 to 300. The 1996 main assessment in mathematics continued to use a cross-grade framework with a 0 to 500 scale. In 1998, reading assessments were scaled across grades, and writing and civics were scaled within-grade.

Linking of scales. Until 2002, results for the main state assessments were linked to the scales for the main national assessments, enabling state and national trends to be studied. Equating the results of the state and national assessments depends on those parts of the main national and state samples that represent a common population: (1) the state comparison sample—students tested in the national assessment who come from the jurisdictions participating in the state NAEP, and (2) the state aggregate sample—the aggregate of all students tested in the state NAEP. Beginning in 2002, the national sample is a subset of the state samples (except in those states that do not participate). Thus no equating is necessary.

Imputation. Up until NAEP's 2002 assessment, no statistical imputations have been generated for missing values in the teacher, school, or SD/LEP questionnaires, nor for missing answers to cognitive questions. Most answers to cognitive questions are missing by design. For example, 8th-grade students being assessed in reading are presented with, on average, 21 out of 110 questions in the assessment. Whether any given student got any of the remaining 89 individual questions right or wrong is not something that NAEP imputes. However, since 1984, multiple imputation techniques have been used to create ***plausible values***. Once created, subsequent users can analyze these plausible values with common software packages to obtain NAEP results that properly account for NAEP's complex item sampling designs.

Because no student takes even a quarter of an assessment, NAEP does not—and cannot—calculate individual scores. Trying to use partial scores based on the small proportion of the assessment to which any given student is exposed would lead to biased results for groups scores due to an inherently large component of measurement error. NAEP developed its process of group score calculation in order to get around the unreliability and noncomparability of NAEP's partial test forms for individuals. NAEP estimates group score distributions using marginal maximum likelihood (MML) estimation, a method that calculates group score distributions based directly on each student's responses to cognitive questions, not on summary scores for each student. As a result, the unreliability of individual-level scores does not decrease NAEP's accuracy in reporting group scores. The MML method does not employ imputations of answers to any questions not of scores for individuals.

NAEP conducts a special form of imputation during the third stage of its analysis procedures. The first stage requires estimating item response theory parameters for each cognitive question. The second stage results in MML estimation of a set of regression coefficients that capture the relationship between group score distributions and nearly all the information from the variables in the teacher, school, or SD/LEP questionnaires, as well as geographical, sample frame, and school record information. The third stage involves calculating imputations designed to reproduce the group-level results that could be obtained during the second stage.

NAEP's imputations follow Rubin's (1987) proposal that the imputation process be carried out several times, so that the variability associated with group score distributions can be accurately represented. NAEP estimates five plausible values for each student. The five plausible values are calculated using the regression coefficients estimated in the second stage. Each plausible value is a random selection from the joint distribution of potential scale scores that fit the observed set of response for each student and the scores for each of the groups to which each student belongs. Estimates based on plausible values are more accurate than if a single (necessarily partial) score were to be estimated for each student and averaged to obtain estimates of subgroup performances. Using the plausible values eliminates the need for secondary analysts to have access to specialized MML software and ensures that the estimates of average performance of groups and estimates of variability in those averages are accurate.

Recent Changes

Several important changes were implemented since 1990. For more detail, refer to earlier sections of this chapter.

- Beginning with the 1990 mathematics assessment, NAGB established three reporting levels for reporting NAEP results: basic, proficient, and advanced.
- In 1990, state assessments were added to NAEP. The 1990 to 1994 assessments are referred to as trial state assessments.
- In 1992, a generalized partial-credit (GPC) model was introduced to develop scales for the more complex constructed-response questions. The GPC model permits the scaling of questions scored according to multipoint rating schemes.
- In 1993, NAGB determined that future NAEP assessments should have within-grade frameworks and scales. The 1994 main history and geography assessments followed this new policy, as did the 1996 main science assessment, the 1996 trend assessments, and the 1998 writing assessment. Mathematics and reading in the main NAEP will continue to have cross-grade scales until further action by NAGB (and a parallel change in the trend assessment).
- In 1994, the new image-based scoring system virtually eliminated paper handling during the scoring process. This system also permits scoring reliability to be monitored online and recalibration methods to be introduced.
- The 1996 main NAEP included new samples for the purpose of studying greater inclusion of SD/LEP students and obtaining data on students eligible for advanced mathematics or science sessions.
- In 1997, there was a probe of student performance in the arts.
- New assessment techniques included: open-ended items in the 1990 mathematics assessment; primary trait, holistic, and writing mechanics scoring procedures in the 1992 writing assessment; the use of calculators in the 1990, 1992, 1996, and 2000 mathematics assessments; a special study on group problem solving in the 1994 history assessment; and a special study in theme blocks in the 1996 mathematics and science assessments.

- In 2001, NAEP fixed the history and geography scales to have within grade scales, with mean of 150, like civics, science, and writing.

- With the expansion of NAEP under the No Child Left Behind Act, NAEP's biennial state-level assessments are being administered by contractor staff (not local teachers). The newly redesigned NAEP has four important features. First, NAEP is administering tests for different subjects (such as mathematics, science, and reading) in the same classroom, thereby simplifying and speeding up sampling, administration, and weighting. Second, NAEP is conducting pilot tests of candidate items for the next assessment 2 years in advance and field tests of items for precalibration 1 year in advance of data collection, thereby speeding up the scaling process. Third, NAEP is conducting bridge studies, administering tests both under the new and the old conditions, thereby providing the possibility of linking old and new findings. Finally, NAEP is adding additional test questions at the upper and lower ends of the difficulty spectrum, thereby increasing NAEP's power to measure performance gaps.

- Beginning with the 2002 assessments, a combined sample of public schools was selected for both state and national NAEP. Therefore, the national sample is a subset of the combined sample of students assessed in each participating state, plus an additional sample from the states that did not participate in the state assessment. This additional sample ensures that the national sample is representative of the total national student population.

- Beginning with the 2003 NAEP, each state must have participation from at least 85 percent—instead of from 70 percent—of the schools in the original sample in order to have results reported.

Future Plans

The next trend assessment will be administered in 2004, and then every 4 years thereafter. For the 21st century, NAEP is undergoing a full-scale redesign, and its assessment schedule is being placed on a more regular, predictable timetable. Main assessments are planned for annual administration (instead of every 2 years). Reading and mathematics will be assessed every 2 years in odd-numbered years; science and writing are planned to be assessed every 4 years (in the same years as reading and mathematics, but alternating with each other); and other subjects will be assessed at the national level in even-numbered years.

5. DATA QUALITY AND COMPARABILITY

As the Nation's Report Card, NAEP must report accurate results for populations of students and subgroups of these populations (e.g., minority students or students attending nonpublic schools). Although only a very small percentage of the student population in each grade is assessed, NAEP estimates are accurate because they depend on the absolute number of students participating, not on the relative proportion of students.

Every activity in NAEP assessments is conducted with rigorous quality control, contributing to both the quality and comparability of the assessments and their results. All questions undergo extensive reviews by subject-area and measurement specialists, as well as careful scrutiny to eliminate any potential bias or lack of sensitivity to particular groups. The complex process by which NAEP data are collected and processed is monitored closely. Although each participating state is responsible for its own data collection for the main state NAEP, Westat ensures uniformity of procedures across states through training, supervision, and quality control monitoring.

With any survey, however, there is the possibility of error. The most likely sources of error in NAEP are described below.

Sampling Error

Two components of uncertainty in NAEP assessments are accounted for in the variability of statistics based on scale scores: (1) the uncertainty due to sampling only a small number of students relative to the whole population, and (2) the uncertainty due to sampling only a relatively small number of questions. The variability of estimates of percentages of students having certain background characteristics or answering a certain cognitive question correctly is accounted for by the first component alone.

Because NAEP uses complex sampling procedures, a jackknife replication procedure is used to estimate standard errors. While the jackknife standard error provides a reasonable measure of uncertainty about student data that can be observed without error, each student in NAEP assessments typically responds to so few questions within any content area that the scale score for the student would be imprecise. It is possible to describe the performance of groups and subgroups of students because as a group all the students are administered a wide range of items.

NAEP uses MML procedures to estimate group distributions of scores. However, the underlying imprecision that makes this step necessary adds an additional component of variability to statistics based on NAEP scale scores. This imprecision is measured by the imputed variance, which is estimated by the variance among the plausible values drawn from each student's posterior distribution of possible scores. The final estimate of the variance is the sum of the sampling variance and the measurement variance.

Nonsampling Error

While there is the possibility of some coverage error in NAEP, the two most likely types of nonsampling error are nonresponse error due to nonparticipation and measurement error due to instrumentation defects (described below). The overall extent of nonsampling error is largely unknown.

Coverage error. In NAEP, coverage error could result from either the sampling frame of schools being incomplete or from the schools' failure to include all the students on the lists from which grade or age samples are drawn. For the 1998 NAEP, the 1997 school list maintained by QED supplied the names of the regular public schools, Bureau of Indian Affairs schools, and DODEA schools. This list, however, did not include schools that opened between 1997 and the time of the 1998 NAEP. To be sure that students in new public schools were represented, each sample district in NAEP was asked to update lists of schools with newly eligible schools.

Catholic and other nonpublic schools were obtained from the NCES Private School Survey (PSS). PSS uses a dual-frame approach. The list frame (containing most private schools in the country) is supplemented by an area frame (containing additional schools identified during a search of randomly selected geographic areas around the country). Coverage of private schools in PSS is very high—estimated at 96.5 percent for the 1995–96 PSS, which was used for the 1998 NAEP. (See chapter 3, section 5.) Prior to the 1996 NAEP, nonpublic schools were also obtained from telephone directories. This process was not repeated in 1996 because the PSS frame adequately supported the QED list.

Nonresponse error.

Unit nonresponse. For both the main NAEP and the trend NAEP, school response rates have generally declined over the years while student response rates have risen. The level of student participation has been consistently lower with each increment in student age and grade. At every age/grade level, the participation of students from nonpublic schools has exceeded that of students from public schools.

For the *main national assessments* in 1998, the unweighted school response rate across grades and subjects was 86 percent (after substitution). This reversed the small declines in national assessment school response rates that occurred between 1990 and 1996. The gains were most likely due to persistent efforts to convert refusals. Between 1990 and 1996, there was a small but steady decline in school response rates despite persistent efforts to convert uninterested schools and districts: from 88.3 to 85.8 percent at grade 4; from 86.7 to 81.9 percent at grade 8; and from 81.3 to 78.7 percent at grade 12. The reason most often given for school nonparticipation is the increase in required testing throughout the jurisdictions and the resulting difficulty in finding time to also conduct NAEP assessments.

Table 10, on the next page, provides weighted response rates for selected NAEP surveys.

Item nonresponse. Specific information about nonresponse for a particular item is available on NAEP summary data tables on the web.

Measurement error. Nonsampling error can result from the failure of the test instruments to measure what is being taught and, in turn, what is being learned by the students. For example, the instruments may contain ambiguous definitions and/or questions that lead to different interpretations by the students. Additional sources of measurement error are the inability or unwillingness of students to give correct information and errors in the recording, coding, or scoring of the data.

To assess the quality of the data in the final NAEP database, survey instruments are selected at random and compared, character by character, with their records in the final database. As in past years, the 2000 NAEP database was found to be more than accurate enough to support analyses. The observed error rates for the 2000 NAEP were comparable to those of past assessments. Error rates ranged from 8 errors per 10,000 responses for the Teacher Questionnaire to 44 errors per 10,000 responses for the School Characteristics and Policies Questionnaire.

Revised results. Following the 1994 assessment, two technical problems were discovered in the procedures used to develop the NAEP mathematics scale and achievement levels determined for the 1990 and 1992

Table 10. Weighted response rates for selected NAEP national (main sample) surveys

		School participation*	Student participation	Overall participation
1994 Reading	– age class 9	86.1	93.5	80.5
	– age class 13	82.9	91.1	75.5
	– age class 17	76.3	81.9	62.5
1996 Mathematics	– grade 4	82.3	95.3	78.4
	– grade 8	81.5	92.9	75.7
	– grade 12	76.2	82.3	62.7
1998 Reading	– grade 4	81.0	96.0	77.8
	– grade 8	76.7	92.7	71.1
	– grade 12	69.7	80.1	55.8

*Rates do not include substitutions.
SOURCE: Allen, Carlson, and Zelenak, *The NAEP 1996 Technical Report* (NCES 1999–452). Allen, Donoghue, and Schoeps, *The NAEP 1998 Technical Report* (NCES 2001–509). Allen, Kline, and Zelenak, *The NAEP 1994 Technical Report* (NCES 97–897).

mathematics assessments. These errors affected the mathematics scale scores reported in 1992 and the achievement level results reported in 1990 and 1992. NCES and NAGB evaluated the impact of these errors and subsequently reanalyzed and reported the revised results from both mathematics assessments. The revised results for 1990 and 1992 are presented in the 1996 mathematics reports. For more detail on these problems, see *NAEP 1996 Technical Report* (NCES 1999–452) and *NAEP 1996 Technical Report of the State Assessment Program in Mathematics* (NCES 97–951).

There were also problems related to reading scale scores and achievement levels. These errors affected the 1992 and 1994 NAEP reading assessment results. The 1992 and 1994 reading data have been reanalyzed and reissued in revised reports. For more information, refer to *The NAEP 1994 Technical Report* (NCES 97–897) and *Technical Report of the NAEP 1994 Trial State Assessment in Reading* (NCES 96–116).

Data Comparability

NAEP allows reliable comparisons between state and national data for any given assessment year. By linking scales across assessments, it is possible to examine short-term trends for data from the main national and state NAEP and long-term trends for data from the long-term trend NAEP.

Main national vs. main state comparisons. NAEP data are collected using a closely monitored and standardized process, which helps ensure the comparability of the results generated from the main national and state assessments. The main national NAEP and main state NAEP use the same assessment booklets, and, beginning in 2002, they are administered in the same sessions using identical procedures.

Short-term trends. Although the test instruments for the main national assessments are designed to be flexible and thus adaptable to changes in curricular and educational approaches, they are kept stable for shorter periods (up to 12 years or more) to allow analysis of short-term trends. For example, through common questions, the 1996 main national assessment in mathematics was linked to both the 1992 and 1994 assessments.

Long-term trends. In order to make long-term comparisons, the long-term trend NAEP uses different samples than the main national NAEP. Unlike the test instruments for the main NAEP, the long-term instruments have remained unchanged from those used in previous assessments. The 1996 trend instruments were identical to those used in the mid-1980s. Through implementation of additional procedures, the current year's data can be linked to even earlier years. The trend NAEP allows the measurement of trends back to 1969, the year of inception. For more detail on the linking of scales in the trend NAEP, refer to section 4, Scaling. The 2004 long-term trend NAEP is undergoing redesign. Bridge studies are planned to make the 2004 assessment comparable to earlier assessments.

Linking to non-NAEP assessments. Linking results from the main state assessments to those from the main national assessments has encouraged efforts to link NAEP assessments with non-NAEP assessments.

Linking to IAEP. In 1992, results from the 1992 NAEP assessments in mathematics were successfully linked to those from the International Assessment of Educational

Progress (IAEP) of 1991. Sample data were collected from U.S. students who had been administered both instruments. The relation between mathematics proficiency in the two assessments was modeled using regression analysis. This model was then used as the basis for projecting IAEP scores from non-U.S. countries onto the NAEP scale. *The relation between the IAEP and NAEP assessments was relatively strong and could be modeled well. The results, however, should be considered only in the context of the similar construction and scoring of the two assessments. Further studies should be initiated cautiously, even though the path to linking assessments is now better understood.*

Linking to TIMSS. The success in linking NAEP to the IAEP sparked an interest in linking the results from the 1996 NAEP assessments in mathematics and science to those from the Third International Mathematics and Science Study (TIMSS) of 1995. The data from this study became available at approximately the same time as the 1996 NAEP data for mathematics and science. Because the two assessments were conducted in different years and no students responded to both assessments, the regression procedure that linked NAEP and IAEP assessments could not be used. The results from grade 8 NAEP and TIMSS assessments were instead linked by matching their distributions. A comparison of the linked results with actual results from states that participated in both assessments suggested that the link was working acceptably. *The results from U.S. students were linked to those of their academic peers in more than 40 other countries. As with the IAEP link, the results should be used cautiously.*

Comparisons with National Adult Literacy Survey (NALS). NAEP data can also be compared with results of NALS. The term "succeed consistently," as it relates to literacy, means that a person at or above a given level of literacy has a certain percentage of a chance of correctly responding to a particular task. The criterion for the NAEP standard (65 percent) is less stringent than the NALS criterion (80 percent). Thus, if the NALS criterion were used for NAEP assessments, the proportions in the lower literacy levels would increase and the proportions in the higher levels would decrease. (See chapter 23 for a description of the NALS.)

Comparisons with IEA Reading Literacy Study. The picture of American students' reading proficiency provided by NAEP assessments is less optimistic than that indicated by the International Association for the Evaluation of Educational Achievement's (IEA) Reading Literacy Study. This can be explained by the following:

(1) The basis for reporting differs considerably between the two assessments. With the IEA, students are compared against other students and not against a standard set of criteria on knowledge, as in NAEP. Much of NAEP reporting is based on comparisons between actual student performance and desired performance (what they are expected to do).

(2) NAEP and IEA assess different aspects of reading. More than 90 percent of the IEA items assess tasks covered in only 17 percent of NAEP items. Further, virtually all of the IEA items are aimed solely at literal comprehension and interpretation, while such items make up only one-third of NAEP reading assessments.

(3) NAEP and IEA differ in what students must do to demonstrate their comprehension. More interpretive and higher level thinking is required to reach the advanced level in NAEP than in the IEA. Also, NAEP requires students to generate answers in their own words much more frequently than does the IEA. Moreover, the IEA test items do not cover the entire expected ability range. Many American students answer every IEA item correctly, making it impossible to distinguish between abilities of students in the upper range. In contrast, the range of item difficulty on NAEP reading assessment exceeds the ability of most American students, so differences in the abilities of students in the upper range can be distinguished easily.

Despite the differences between these two assessments, there is a high probability that, if students from other countries were to take NAEP, the rank ordering or relative performance of countries would be about the same as in the IEA findings. This assumption is based on the theoretic underpinnings of item response theory and its application to the test scaling used for both the IEA Reading Literacy Study and the NAEP reading assessment.

6. CONTACT INFORMATION

For content information on NAEP, contact:

Peggy Carr
Phone: (202) 502-7321
E-mail: peggy.carr@ed.gov

Steven Gorman
Phone: (202) 502-7347
E-mail: steven.gorman@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

The NAEP Guide: A Description of the Content and Methods of the 1997 and 1998 Assessments, NCES 97–990, by J. Calderone, L.M. King, and N. Horkay, eds. Washington, DC: 1997.

The NAEP 1998 Technical Report, NCES 2001–509, by N.L. Allen, J.R. Donoghue, and T.L. Schoeps. Washington, DC: 2001.

The NAEP 1994 Technical Report, NCES 97–897, by N.L. Allen, D.L. Kline, and C.A. Zelenak. Washington, DC: 1996.

The NAEP 1996 Technical Report, NCES 1999–452, by N.L. Allen, J.E. Carlson, and C.A. Zelenak. Washington, DC: 1999.

NAEP 1992 Technical Report, NCES 94–490, by E.G. Johnson and J.E. Carlson. Washington, DC: 1994.

The 1998 High School Transcript Study User's Guide and Technical Report, NCES 2001–477, by S. Roey, N. Caldwell, K. Rust, E. Blumstein, T. Krenzke, S. Legum, J. Kuhn, M. Waksberg. Washington, DC: 2001.

Overview of NAEP Assessment Frameworks, NCES 94–412, by S. White. Washington, DC: 1994.

Procedures Guide for Transcript Studies, NCES Working Paper 1999–05, by M.N. Alt and D. Bradby. Washington, DC: 1999.

Technical Issues in Large-Scale Performance Assessment, NCES 96–802, by G. Phillips and A. Goldstein. Washington, DC: 1996.

Technical Report of the NAEP 1996 State Assessment Program in Mathematics, NCES 97–951, by N.L. Allen, F. Jenkins, E. Kulick, and C.A. Zelenak. Washington, DC: 1997.

Technical Report of the NAEP 1992 Trial State Assessment in Reading, NCES 94–472, by E.G. Johnson, J. Mazzeo, and D.L. Kline. Washington, DC: 1994.

Technical Report of the NAEP 1994 Trial State Assessment in Reading, NCES 96–116, by J. Mazzeo, N.L. Allen, and D.L. Kline. Washington, DC: 1995.

Technical Report: NAEP 1996 State Assessment Program in Science, NCES 98–480, by N.L. Allen, S.S. Swinton, S.P. Isham, and C.A. Zelenak. Washington, DC: 1998.

Uses of Data

Focus on NAEP: New Software Makes NAEP Data User-Friendly, NCES 97–045, by A. Vanneman. Washington, DC: 1997.

Interpreting NAEP Scales, NCES 93–421, by G. Phillips. Washington, DC: 1993.

Survey Design

ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results, NCES Working Paper 97–39, by L. Bay, L. Chen, B.A. Hanson, J. Happel, M.J. Kolen, T. Miller, M. Pommerich, J. Sconing, T. Wang, and C. Welch. Washington, DC: 1997.

Can State Assessment Data Be Used to Reduce State NAEP Sample Sizes?, NCES Working Paper 97–29, by D. McLaughlin. Washington, DC: 1997.

Exploring New Methods for Collecting Students' School-Based Writing, NCES 92–065, by E.H. Owen. Washington, DC: 1992.

The Inclusion of Students with Disabilities and Limited English Proficient Students in Large-Scale Assessments: A Summary of Recent Progress, NCES 97–482, by J.F. Olson and A.A. Goldstein. Washington, DC: 1997.

Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires), NCES Working Paper 97–32, by R.G. Niemi. Washington, DC: 1997.

Multiple Imputation for Nonresponse in Surveys, by D.B. Rubin. New York: John Wiley & Sons, 1987.

NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress, NCES Working Paper 97–31, by E.G. Johnson, S. Lazer, and C.Y. O'Sullivan. Washington, DC: 1997.

Optimal Rating Procedures and Methodology for NAEP Open-ended Items, NCES Working Paper 97–37, by R.J. Patz, M. Wilson, and M. Hoskens. Washington, DC: 1997.

Data Quality and Comparability

Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress, by J.W. Pellegrino, L.R. Jones, and K.J. Mitchell, eds. National Research Council. Washington, DC: 1999.

Grading the Nation's Report Card: Research from the Evaluation of NAEP, by N.S. Raju, J.W. Pellegrino, M.W. Bertenthal, K.J. Mitchell, and L.R. Jones, eds. National Research Council. Washington, DC: 2000.

Linking the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS): A Technical Report, NCES 98–499, by E.G. Johnson. Washington, DC: 1998.

Model-Based Methods for Analysis of Data from 1990 NAEP Trial State Assessment, NCES 95–696, by D.A. Sedlacek. Washington, DC: 1995.

NAEP Reporting Practices: Investigating District-Level and Market-Basket Reporting, by P.J. DeVito and J.A. Koenig, Editors. National Research Council. Washington, DC: 2001.

National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data, NCES Working Paper 95–06, by S. Ingels and J. Taylor. Washington, DC: 1995.

Use of Person-Fit Statistics in Reporting and Analyzing National Assessment of Educational Progress Results, NCES 95–713, by D.A. Sedlacek. Washington, DC: 1995.

Using HLM and NAEP Data to Explore School Correlates of 1990 Mathematics and Geometry Achievement in Grades 4, 8, 12—Methodology and Results, NCES 95–697, by D.A. Sedlacek. Washington, DC: 1995.

Chapter 21: Third International Mathematics and Science Study (TIMSS)

1. OVERVIEW

The Third International Mathematics and Science Study (TIMSS), sponsored by the International Association for the Evaluation of Educational Achievement (IEA), is a study of classrooms across the country and around the world. A half million students from 41 countries were tested in 30 different languages at five different grade levels to compare their mathematics and science achievement. Intensive studies of students, teachers, schools, curriculum, instruction, and policy issues were also carried out to understand the educational context in which learning takes place.

TIMSS represents the continuation of a long series of studies conducted by the IEA. The IEA conducted its First International Mathematics Study (FIMS) in 1964 and the Second International Mathematics Study (SIMS) in 1980–82. The First and Second International Science Studies (FISS and SISS) were carried out in 1970–71 and 1983–84, respectively. Since the subjects of mathematics and sciences are related in many respects and since there is broad interest among countries in students' abilities in both mathematics and science, the third studies (TIMSS) were conducted as an integrated effort.

TIMSS collected data from students in three separate populations. *Population 1*, in which 26 countries participated, consisted of students enrolled in the two adjacent grades that contained the largest proportion of 9-year-old students at the time of testing; in most countries, these were the 3rd and 4th grades. *Population 2*, in which 41 countries participated, consisted of students enrolled in the two adjacent grades that contained the highest proportion of 13-year-old students at the time of testing; in most countries, these were the 7th and 8th grades. *Population 3*, in which 23 countries participated, consisted of students in their final year of secondary education. As an additional option, countries could test special subgroups of these students: students having taken advanced courses in mathematics and students having taken courses in physics.

In 1999, a follow-up study called the Third International Mathematics and Science Study-Repeat (TIMSS-R) was conducted. The design of TIMSS-R makes it possible to track changes in achievement and certain background factors from the first TIMSS study. It incorporated an expanded videotape classroom study as well as a National Assessment of Educational Progress (NAEP)/TIMSS linking study to allow researchers to compare TIMSS results with those from NAEP. In addition, the TIMSS-R included a national Benchmarking Project, through which districts and states in the United States could compare their progress internationally as individual "nations." Unlike the first TIMSS, the 1999 TIMSS-R study focused only on 8th-grade students.

1995 WORLDWIDE STUDY OF CLASSROOMS, WITH FOLLOW-UP STUDY OF 8th GRADERS IN 1999; 41 COUNTRIES PARTICIPATED

TIMSS tested three populations:

- 9-Year-Olds/3rd and 4th Graders
- 13-Year-Olds/7th and 8th Graders
- Students in Final Year of Secondary Education

BEST COPY AVAILABLE

Purpose

The two broad questions that TIMSS addresses are: (1) How do mathematics and science educational environments differ across countries, how do student outcomes differ, and how are differences in those outcomes related to differences in mathematics and science education environments? (2) Are there patterns of relationships among contexts, inputs, and outcomes within countries that can lead to improvements in the theories and practices of mathematics and science education?

Components

TIMSS used several types of instruments to collect data about students, teachers, and schools. In addition, 8th graders in the United States, Japan, and Germany participated in a ***videotape study***, in which actual classroom sessions were recorded, coded, and analyzed; this study was expanded to include seven nations in TIMSS-R. Various populations also participated in ***curriculum studies*** and ***ethnographic case studies***. The United States sponsored two additional components of TIMSS-R: a Benchmarking Project and the NAEP/TIMSS-R Linking Study. The TIMSS-R ***did not*** include the performance assessment.

Written Assessment. Questionnaires were developed to test Population 1, 2, and 3 students in various content areas within mathematics and science. For Population 1, the mathematics content areas included: whole numbers; fractions and proportionality; measurement, estimation, and number sense; data representation, analysis, and probability; geometry; and patterns, relations, and functions. The Population 1 science content areas were earth science; life science; physical science; and environmental issues and the nature of science. The Population 2 mathematics content areas were fractions and number sense; geometry; algebra; data representation, analysis, and probability; measurement; and proportionality. The Population 2 science content areas were earth science; life science; physics; chemistry; and environmental issues and the nature of science. The Population 3 mathematics contents areas were numbers; measurement; geometry; proportionality; functions, relations, and equations; data, probability, and statistics; elementary analysis; and validation and structure. The Population 3 science contents were earth sciences; life sciences; physical sciences; science, technology, and mathematics; history of science; environmental issues; nature of science; and science and other disciplines. In addition, Population 3 students who had taken advanced mathematics were eligible for the advanced mathematics test, which included numbers and equations, calculus, geometry, probability and statistics, and validation and structure. Population 3 students who had taken physics were eligible for a physics test. Its contents were mechanics, electricity and magnetism, heat, wave phenomena, and modern physics—particle, quantum and astrophysics, and relativity.

TIMSS-R written assessment tests repeat the Population 2 content areas.

Student Background Questionnaire. The student questionnaire for Populations 1 and 2 asked about students' demographics and home environment, including academic activities outside of school, people living in the home, parental education (only at Population 2), books in the home, possessions in the home, and the importance of students' mothers, peers, and friends placed on different aspects of education. Students were also queried about their attitudes toward mathematics and science. The final sections of the questionnaires asked about classroom experiences in mathematics and science. Similar items were asked of students in TIMSS-R.

The student questionnaire administered to Population 3 students was similar in most respects to the Population 2 student questionnaires. The only differences were that Population 3 students were also queried as to their future plans, their programs of study, and the most advanced mathematics and science courses they had taken.

Teacher Questionnaire. The teacher questionnaires for Population 2 addressed four major areas: teachers' background, instructional practices, students' opportunity to learn, and teachers' pedagogic beliefs. There are separate questionnaires for teachers of mathematics and of science. Since most Population 1 teachers teach all subjects, a single teacher questionnaire was developed to address both mathematics and science. So as not to overburden the teachers, the classroom practice questions in the Population 1 teacher questionnaire pertain mostly to mathematics. However, teachers also were asked about how they spend their time in school and the atmosphere in their schools (e.g., teaching loads, collaboration policies, responsibilities for decision-making, and the availability of resources).

The teacher questionnaires were designed to provide information about the teachers of the student samples in TIMSS. The teachers who completed TIMSS questionnaires do not constitute a sample from any definable population of teachers. Rather, they represent the teachers of a national sample of students.

BEST COPY AVAILABLE

There was no teacher questionnaire administered to the teachers of students in Population 3.

The teacher questionnaire for TIMSS-R gathered data about topics such as attitudes and beliefs about teaching and learning, teaching assignments, class size and organization, topics covered, the use of various teaching tools, instructional practices, and participation in professional development.

School Questionnaire. The school questionnaires for each population sought information about the school's community, staff, students, curriculum and programs of study, and instructional resources and time. At Populations 1 and 2, the school questionnaires also ask about the number of years students are taught by the same teacher. A school questionnaire was to be completed by the principal, headmaster, or other administrator of each school that participated in TIMSS. Similar items were asked of principals in TIMSS-R.

Performance Assessment. The TIMSS performance assessment was administered at Populations 1 and 2 to a subsample of students in the upper grades that participated in the written assessment. The performance tasks permitted students to demonstrate their ability to make, record, and communicate observations; to take measurements or collect experimental data and present them systematically; to design and conduct a scientific investigation; or to solve certain types of problems. A set of 13 such hands-on activities was developed; 11 of these tasks were either identical or similar across populations, and 2 were different. Of these two, one task was administered to Population 1 (4th graders) and one was administered to Population 2 (8th graders).

Videotape Study. The videotape classroom study was designed as the first study to collect videotaped records of classroom instruction from national probability samples in Japan, Germany, and the United States to gather more in-depth information about the context in which learning takes place and also to enhance understanding of the statistical indicators available from the main TIMSS study. An hour of regular classroom instruction was videotaped in a subsample of 8th-grade mathematics classrooms (except in Japan, where videotaping was usually done in a different class, selected by the principal) included in the assessment phase of TIMSS in each of the three countries.

National-level univariate statistics were constructed to generate descriptive statistics for each country and a comparison was made between the mathematics achievement scores of classrooms in the main TIMSS samples and the subsample of classrooms selected for the video study.

The TIMSS-R Videotape Classroom Study was expanded in scope to examine national samples of 8th-grade mathematics and science instructional practices in seven nations: Australia, the Czech Republic, Hong Kong, Japan, the Netherlands, Switzerland, and the United States. Four countries—Australia, the Czech Republic, the Netherlands, and the United States—participated in both the mathematics and science components of the study. Hong Kong and Switzerland participated in only the mathematics component, and Japan in only the science component.

Curriculum Studies. Continuing the approach of previous IEA studies, TIMSS addressed three conceptual levels of curriculum. The intended curriculum is composed of the mathematics and science instructional and learning goals as defined at the system level. The implemented curriculum is the mathematics and science curriculum as interpreted by teachers and made available to teachers. The attained curriculum is the mathematics and science content that students have learned and their attitudes toward these subjects. To aid in interpretation and comparison of results, TIMSS also collected extensive information about the social and cultural contexts for learning, many of which are related to variation among educational systems.

To gather information about the intended curriculum, mathematics and science specialists within each participating country worked section by section through curriculum guides, textbooks, and other curricular materials to categorize aspects of these materials in accordance with detailed specification derived from TIMSS mathematics and science curriculum frameworks.

To collect data about how the curriculum is implemented in classrooms, TIMSS administered a broad array of questionnaires, which also collected information about the social and cultural contexts for learning. Questionnaires were administered at the country level about decision-making and organizational features within the education systems. The students who were tested answered questions pertaining to their attitudes toward mathematics and science, classroom activities, home background, and out-of-school activities. The mathematics and sciences teachers of sampled students responded to questions about teaching emphasis on the topics in the curriculum frameworks, instructional practices, textbook use, professional training and education, and their views on mathematics and science. The heads of schools responded to questions about school staffing and resources, mathematics and science course offerings, and support for teachers. In addition a volume was complied that presents descriptions of the educational systems of the participating countries.

Ethnographic Case Studies. The case studies approach to understanding cultural differences in behavior has a long history in selected social science fields. Given the goals of TIMSS, it was designed to focus on four key topics that challenge U.S. policymakers and investigate how these topics are dealt with in the United States, Japan, and Germany: implementation of national standards; the working environment and training of teachers; methods for dealing with differences in ability; and the role of school in adolescents' lives. Each topic was studied through interviews with a broad spectrum of students, parents, teachers, and educational specialists. The ethnographic approach permitted researchers to explore the topics in a naturalistic manner and to pursue them in greater or lesser detail, depending on the course of the discussion. As such, these studies both validate and integrate the information gained from official sources with that obtained from teachers, students, and parents in order to ascertain the degree to which official policy reflects actual practice. The objective is to describe policies and practices in the nations under study that are similar to, different from, or nonexistent in the United States.

In three regions in each of the three countries, the research plan called for each of the four topics to be studied in the 4th, 8th, and 12th grades. The specific cities and schools were selected "purposively" to represent different geographical regions, policy environments, and ethnic and socioeconomic backgrounds. Schools in the case studies were separate from schools in the main TIMSS sample. Where possible, a shortened form of the TIMSS test was administered to the students in the selected schools. The ethnographic researchers in each of the countries conducted interviews and obtained information through observations in schools and homes. Both native-born and nonnative researchers participated in the study to ensure a range of perspectives.

TIMSS-R Benchmarking Project. Twenty-seven states, districts and consortia of districts throughout the United States participated as their own "nations" in this project, following the same guidelines as the participating countries. The samples drawn for each of these states and districts are representative of the student population in each of these states and districts. The findings from this project allow these jurisdictions to assess their comparative international standing and judge their mathematics and science programs in an international context.

NAEP/TIMSS-R Linking Study. A subsample of students taking the 2000 state NAEP mathematics and science assessment also took the TIMSS-R assessment. (See chapter 20 for more information on NAEP.) This provides an opportunity to compare students' performance on NAEP to their performance on TIMSS-R, and allows for estimates of how states participating in NAEP 2000 would have performed had they participated in TIMSS-R. Results from the TIMSS-R Benchmarking Study are used to check the results of the linking study.

Periodicity

The Third International Mathematics and Science Study was conducted only once. Previous international math studies were conducted in 1964 and 1980–82; previous international science studies were conducted in 1970–71 and 1983–84. A follow-up study of 8th graders, using a similar design (but different students) was conducted in 1999. This follow-up study is called the Third International Mathematics and Science Study-Repeat (TIMSS-R).

2. USES OF DATA

The possibilities for specific research questions to be dealt with by TIMSS are numerous; however, the main research questions, focused at the student, the school or classroom, and the national or international levels, are illustrated below:

- How much mathematics and science have students learned?
- How well are students able to apply mathematics and science in problem-solving abilities?
- What are students' attitudes toward mathematics and science?
- How do gender differences in participation rates, course selection, and student outcomes differ across countries?
- What do teachers teach in their classrooms?
- What methods and materials do teachers use in teaching mathematics and science, and how are they related to student outcomes?
- What kinds of grouping practices, either within or between classrooms, are used, and how are those practices reflected in student outcomes and participation in subsequent mathematics and science courses?

- How strongly are students motivated to learning in general and to the learning of mathematics and science in particular? What are the sources of their motivation?

- What factors characterize the academic and professional preparation of teachers of mathematics and science?

- What are teachers' beliefs and opinions about the nature of mathematics and science and their teaching, and how are these related to comparable opinions and attitudes of their students?

- How do teachers evaluate their students?

- If there are national curricula in a country, how specific are they, and what efforts are made to see that the national curricula are followed?

- What proportions of students plan to study mathematics or science at the postsecondary level or to pursue mathematics or science-based careers?

Country-level outcomes are necessarily related to student- and classroom-level outcomes, and an important aspect of TIMSS is to identify the prime determinants of student outcomes, including the amount and quality of opportunity to learn and the intensity and perseverance of the students' motivation.

3. KEY CONCEPTS

Key terms related to TIMSS are described below.

Nationally Desired Population. The effective target population within each participating country. The stated objective in TIMSS was that the Nationally Desired Population within each country be as close as possible to the International Desired Population, which is the target population. (See below.) Using the International Desired Population as a basis, participating countries had to operationally define their populations for sampling purposes. Some National Research Coordinators had to restrict coverage at the county level, for example, by excluding remote regions or a segment of the educational system. Thus, the Nationally Desired Population sometimes differed from the International Desired Population.

National Research Coordinators (NRCs). The official from each participating country appointed to implement national data collection and processing in accordance with international standards. In addition to selecting the sample of students to be taken, NRCs were responsible for working with school coordinators, translating the test instruments, assembling and printing the test booklets, and packing and shipping the necessary materials to the sampled schools. They were also responsible for arranging the return of the testing materials from the school to the national center, preparing for and implementing the free-response scoring, entering the results into data files, conducting on-site quality assurance observations for a 10 percent sample of schools, and preparing a report on survey activities.

4. SURVEY DESIGN

Target Population

For TIMSS Populations 1 and 2, the International Desired Populations for all countries were defined as follows:

- Population 1: All students enrolled in the two adjacent grades that contain the largest proportion of 9-year-olds at the time of testing

- Population 2: All students enrolled in the two adjacent grades that contain the largest proportion of 13-year-olds at the time of testing

TIMSS used a grade-based definition of the target population at Populations 1 and 2. In a few cases, TIMSS components were administered only to the upper grade of these populations (i.e., the performance assessment was conducted at the upper grade and some background questions were asked of the upper grade students only). However, two adjacent grades were chosen to ensure extensive coverage of the same age cohort for most countries, thereby increasing the likelihood of producing useful age-based comparisons in addition to the grade-based analyses.

The intention of the assessment of final-year students (Population 3) was to measure what might be considered the "yield" of the elementary and secondary education systems of a country with regard to mathematics and science. This was accomplished by assessing the mathematics and science literacy of all students in the final year of secondary school, the advanced mathematics knowledge of students having taken advanced mathematics courses, and the physics knowledge of students having taken physics. The International Desired Population, then, was all students in the final year of secondary school, with students having taken advanced mathematics courses and students having taken physics courses as two overlapping subpopulations. Students repeating the final year were not part of the desired population. For each secondary education track in a country, the final grade of the track was identified as being part of the target population, allowing substantial coverage of students in their final year of schooling. For example, grade 10 could be

BEST COPY AVAILABLE

the final year of a vocational program, and grade 12 the final year of an academic program. Both of these grade/track combinations are considered part of the target population, but grade 10 in the academic track is not.

For TIMSS-R, the international desired population consisted of all students in each participating nations who were enrolled in the upper of the two adjacent grades that contained the greatest proportion of 13-year-olds at the time of testing.

Sample Design

The TIMSS sample design for each country and population was intended to give a probability sample of all students within the target grades in the national school system (except for a small number of students allowed to be excluded as ineligible according to national criteria). Every eligible student in the country's school system had a chance of being selected, with a fixed probability of selection. These probabilities of selection were designed to be equal across eligible students as much as was possible, but for a variety of reasons the eligible students' probabilities of selection differ between students in most of the national samples.

Written Assessment. The TIMSS sample design was a two-stage cluster sample, with schools as the first stage of selection and classrooms within schools as the second stage of selection. The classroom sampling design was intended to be an equal probability design with no subsampling in the classroom. However, a design based on a probability proportionate to size (PPS) sample of classrooms with a fixed sample size of students selected within the sampled classroom was permitted under the international guidelines. Exclusions could occur at the school level, the student level, or both. TIMSS participants were expected to keep such exclusions to no more than 10 percent of the national desired population. Twenty of 23 participants in the Population 3 study achieved 100 percent coverage. The school sampling process was generally a stratified probability PPS sample, with the measure of size for a school equal to the number of students in the school in the two target grades for each population.

In the first stage of sampling, representative samples of schools were selected from sampling frames (comprehensive lists of all eligible students). TIMSS standards for sampling precision required that all population samples have an effective sample size of at least 400 students for the main criterion variables. To meet the standard, at least 150 schools were to be selected per target population. However, the clustering effect of sampling classrooms rather than students was also considered in determining the overall sample size for TIMSS. Because the magnitude of the clustering effect is determined by the size of the cluster and the intraclass correlation, TIMSS produced sample-design tables showing the number of schools to sample for a range of intraclass correlations and minimum-cluster-size values. Some countries needed to sample more than 150 schools. Countries, however, were asked to sample 150 schools even if the estimated number of schools to sample was less than 150.

The schools in each explicit stratum (e.g., geographical region, public/private, etc.) were listed in order of the implicit stratification variables, and then further sorted according to their measure of size. Of course, the stratification variables differed from country to country. Small schools were handled either through explicit stratification or through the use of pseudo-schools. In some very large countries, there was a preliminary sampling stage before schools were sampled in which the country was divided into primary sampling units.

In cases where a sampled school was unable to participate in the assessment, it was replaced by a replacement school. The mechanism for selecting replacement schools, established a priori, identified the next school on the ordered school-sampling list as the replacement for each particular sampled school. The school after that was a second replacement, should it be necessary. Using either explicit or implicit stratification variables and ordering of the school sampling frame by size ensured that any original sampled school's replacement would have similar characteristics.

In the second sampling stage, classrooms of students were sampled. Generally, in each school, one classroom was sampled from each target grade, although some countries opted to sample two classrooms at the upper grade in order to be able to conduct special analyses. Most participants tested all students in selected classrooms, and in these instances the classrooms were selected with equal probabilities. A few participants used a design based on a PPS sample of classrooms, with a fixed sample size of students selected within the sampled classrooms.

In an optional third sampling stage, participants with particularly large classrooms in their schools could decide to subsample a fixed number of students from each selected classroom. This was done using a simple random sampling method whereby all students in a sampled classroom were assigned equal selection probabilities.

BEST COPY AVAILABLE

For Population 3, in order to implement the TIMSS goal of assessing the mathematics and science literacy of all students while also assessing the advanced mathematics and physics knowledge of students with preparation in these subjects, it was necessary to develop a sampling design that ensured that students were stratified according to their level of preparation in mathematics and physics, so that appropriate test booklets could be assigned to them. Within each sampled school, students were classified according to a four-group classification scheme (i.e., students having studied neither advanced mathematics nor physics, students having studied physics but not advanced mathematics, students having studied advanced mathematics but not physics, and students having studied both advanced mathematics and physics), and 40 students were sampled at random, 10 from each of the four categories. If just three student types were present three samples of 13 students were drawn. In some tracked systems, schools frequently consisted of a single group. In these situations all 40 students were sampled from whichever group was appropriate.

The United States' national TIMSS design followed the international specifications described above for the three populations. Primary sampling units (PSUs) were sampled as the first stage of sampling with the PSUs defined as metropolitan statistical areas (MSAs), single counties, or groups of counties. There were 1,027 PSUs on the sampling frame with 11 of the PSUs taken as certainty selections (representing the 11 largest metropolitan areas) and 48 PSUs drawn from the remaining 1,016 PSUs, with probability proportionate to the 1990 population within the PSU. These PSUs were placed in eight primary strata. The 48 noncertainty PSUs were substratified by socioeconomic status and demographic characteristics that were found to be most highly related to educational achievement within the primary strata, as measured by aggregated assessment data from previous NAEP surveys. (For more information on NAEP, see chapter 20.)

For both the 11 certainty PSUs and the 48 sampled noncertainty PSUs, the measures of the size of the school were proportional to the target grade size in the school divided by the PSU probability of selection. In addition, schools in both types of PSUs with high percentages of Blacks, and Hispanics (greater than 15 percent of the population) were given doubled probabilities of selection. The school sample sizes for both Populations 1 and 2 were 220 schools.

Public and private schools were sampled from separate frames. The public school sample was drawn from the most recent Quality Education Data (QED) sampling frame. The private schools sample was drawn from the 1991–1992 Private School Universe Survey (PSS) file. (For more information on PSS, see chapter 3.)

The U.S. sample design within schools for Populations 1 and 2 consisted of an equal probability sample of two upper grade (4th- or 8th-grade) classrooms and one lower grade (3rd- or 7th-grade) classroom within the school. All eligible students in the classroom were designated to be in the sample (i.e., there was no subsampling of students in the U.S. sample). The extra sampled classroom in the upper grade beyond the international minimum was drawn for the purpose of permitting analyses that did not confound school effects and classroom effects for grades 4 and 8. Classrooms were sampled with equal probability for each target grade in each sampled school in the U.S. sample, in accord with international specifications. All students in the sampled classroom were taken in the TIMSS sample. The sample design was approximately self-weighting at the student level within particular subgroups of the schools.

Performance Assessment. For the performance assessment, TIMSS participants were to sample at least 50 schools from those already selected for the written assessment, and from each school a sample of either 9 or 18 upper-grade students already selected for the written assessment. This yielded a sample of about 450 students in the upper grade of Populations 1 and 2 (4th and 8th grades in most countries) in each country. For the performance assessment, in the interest of ensuring the quality of administration, countries could exclude additional schools if the schools had fewer than nine students in the upper grade or if the schools were in a remote region. The exclusion rate for the performance assessment was not to exceed 25 percent of the national desired population.

Teacher Questionnaire. The TIMSS database for each country includes questionnaire data from the teachers of the sampled classrooms, which can be linked to student assessment data in the classrooms. Any teacher linked as a mathematics or science teacher to any assessed student is eligible to receive a questionnaire. The classroom sample is drawn from a listing of mathematics classrooms, so that in most situations only one mathematics teacher is linked to each sampled classroom. If this single teacher is also only linked to single sampled classroom, then the teacher received a questionnaire for that single classroom.

This straightforward one-to-one linking does not always hold, however. In some cases, teachers may teach both

BEST COPY AVAILABLE

mathematics and science to students in a sampled classroom, making them eligible to receive questionnaires for both subjects. For a single subject, a teacher may also teach multiple classrooms (e.g., the sampled classrooms for the school from both target grades).

For the U.S. TIMSS sample, a teacher was never asked to complete more than one questionnaire. In cases when a teacher taught both subject areas, the teacher was randomly assigned to receive a mathematics or science teachers' questionnaire. In cases when a teacher taught assessed students in one subject area in more than one classroom, the teacher was purposively assigned one classroom.

Each country was allowed to develop its own methodology for this process of assigning subjects and classrooms to teachers when the links were not straightforward due to the presence of one to many (or many to one) mappings.

Videotape Study. The sample for the TIMSS videotape study was assembled as a subsample of Population 2 students in Germany, Japan, and the United States. In the United States, schools were selected for the video study as follows: First, Population 2 TIMSS schools were listed in the order in which they were originally sampled. Using this ordering, pairs of schools were generated. Within each pair one of the two schools was randomly sampled (with each school having an equal probability of being sampled). The unsampled school in the pair was reserved as a potential replacement for the sampled school. A total of 109 pairs were assigned, with one school unpaired because one school of the original Population 2 sample of 220 schools had no 8th grade. The final videotape study sample size was 109. The unpaired school was not sampled. Within each sampled school, one 8th-grade classroom was selected with equal probability from the two TIMSS 8th-grade classrooms in the school. There was no sorting or stratification of classrooms by level of mathematics taught. In the event that the sampled teacher refused to be videotaped, the classroom was never replaced by the other 8th-grade classroom in the same school. Instead the entire school was replaced by its paired school.

The final TIMSS video sample in the United States consisted of 81 schools, of which 73 were public schools and 8 were private schools. The final video sample in Germany consisted of 100 schools, 15 of which were replacement schools. In Japan, 50 schools participated in the videotape study, 2 of which were replacement schools.

Sampling for the TIMSS-R videotape study was performed in two steps. The first step was to sample 100 schools in each country. The second step was to sample one mathematics classroom and one science classroom from each school. Sampling of schools in each country was performed using the same procedures being used in the TIMSS-R achievement study; most countries, however, did not videotape in the same schools in which the TIMSS-R assessment was conducted. Thus, linkage of the video study to the achievement study is only at the national level. A replacement school will be chosen for each of the 100 schools for each country. If the primary school refused to participate, its replacement school was invited to replace it. Within each school, one mathematics class and/or one science class was randomly selected for videotaping.

Assessment Design

The task of putting together the achievement item pools for the different TIMSS tests took more than 3 years to complete. The process necessitated building international consensus among NRCs, their national committees, mathematics and science experts, and measurement specialists. The NRCs from all participating countries worked to ensure that the items used in the tests were appropriate for their students and reflected their country's curricula. Because students in Population 3 were less likely to have been taught a comparable curriculum (due to some students' having taken advanced mathematics and physics classes), the design of written assessments for this population differs somewhat from that of Populations 1 and 2. As a result, Population 3 will be discussed separately.

The international versions of the test instruments and the student and school background questionnaires were developed in English and then translated into other languages by TIMSS countries. While the intent of TIMSS was to provide internationally comparable data for all variables, there were many contextual differences among countries so that the international version of the questions was not always appropriate in all countries. Therefore, the international versions of the questionnaires were designed to provide an opportunity for individual countries to modify some questions or response options in order to include the appropriate wording or options most consistent with their own national systems. Each item deviation or national adaptation was reviewed to determine whether the national data should be: deleted as not being internationally comparable, recoded to match the international version, or retained with some documentation describing modifications. Whenever possible, national data were retained to match as closely as possible the international version of the items and/or by documenting minor deviations.

BEST COPY AVAILABLE

For Populations 1 and 2, the test items were allocated to 26 different clusters. Also, at each population, the 26 clusters were assembled into eight booklets. Each student completed one booklet. At Population 1, the clusters were either 9 or 10 minutes in length. The core cluster, which was composed of five mathematics and five science multiple-choice items, was included in all booklets. Focus clusters appeared in at least three booklets, so that the items were answered by a relatively large fraction (three-eighths) of the student sample in each country. The breadth clusters, largely containing multiple-choice items, appeared in only one booklet each. The free-response clusters were each assigned to two booklets, so that items statistics of reasonable accuracy would be available. The booklet design for Population 2 is very similar to that for Population 1, differing only in the length and item content of the clusters.

Students in Population 3 were classified into four groups based on their preparation in mathematics and physics. Each student was characterized as having taken advanced mathematics or not, and as having taken physics or not. The assessment of these students was accomplished through a complex design that included four types of test booklets (nine booklets in total) that were distributed to students based on their academic preparation. The four types of test booklets were intended to yield proficiency estimates in mathematics and science literacy, advanced mathematics, and physics.

The TIMSS test design for Population 3 included 12 mutually exclusive clusters of items distributed among the four types of test booklets in a systematic fashion. The test booklets were rotated among students based on the student classification scheme so that each student completed one 90-minute test booklet.

TIMSS-R utilized the same assessment framework designed for TIMSS. Approximately one-third of the original 1995 TIMSS assessment items were kept secure so that they could be included in the 1999 TIMSS-R assessment to provide trend data. For the approximately two-thirds of items that were released to the public, a panel of international assessment and content experts and the NRCs of each participating country developed and reviewed replacement items that closely matched the content of the original items. Through this process, over 300 science and mathematics items were developed as potential replacement items, of which 277 items were carefully chosen to be field tested. Approximately 1,000 students per country participated in this field test. Of the 277 potential replacement items, 202 were selected based on the results of the field test.

Data Collection and Processing

Each country participating in TIMSS was responsible for collecting its national data and processing the materials in accordance with the international standards. In each country, a national research center and NRC were appointed to implement these activities. One of the main ways in which TIMSS sought to achieve uniform project implementation was by providing clear and explicit instructions on all operational procedures. Such instructions were provided primarily in the form of operations manuals, supported where possible by computer software systems that assisted NRCs in carrying out the specified filed operations procedures. Forms accompanying some of the manuals served to document the implementation of the procedures in each country. Many of these forms were used to track schools, students, and teachers, and to ensure proper linkage of schools, students, and teachers in the database.

Reference dates. All TIMSS testing was conducted at "the end of the school year." Because academic schedules differ across countries, this was not a set date for all countries, but was relative to each country's particular educational system. Most countries tested the mathematics and science achievement of their students at the end of the 1994–95 school year, most often in May and June of 1995. The three countries on a Southern Hemisphere school schedule (Australia, New Zealand, and South Africa) tested between August and December 1995, which was late in the school year in the Southern Hemisphere. Three countries (Iceland, Germany, and Lithuania) tested their final-year students (or a subset of them) at the end of the 1995–96 school year.

Likewise, TIMSS-R was conducted on two schedules. The Southern Hemisphere countries administered the survey from September to November, 1998, while the Northern Hemisphere countries did so from February to May, 1999.

Data collection. Each participating country was responsible for carrying out all aspects of the data collection, using standardized procedures developed for the study. Training manuals were created for school coordinators and test administrators that explained procedures for receipt and distribution of materials as well as for the activities related to the training sessions. The manuals covered procedures for test security, standardized scripts to regulate directions and timing, rules for answering students' questions, and steps to ensure that identification on the test booklets and questionnaires corresponded to the information on the forms used to track students.

Specific discussions of collection methods for the performance assessment and videotape study are provided below.

Performance Assessment. Specific procedures were established to ensure that the performance assessment was administered in as standardized a manner as possible across countries and schools. The NRC in each participating country was responsible for collecting the equipment and materials required for each of the performance assessment tasks, and for assembling a set of materials for each school. The tasks were designed to require only materials that were easy to obtain and inexpensive. Many of the pieces of "equipment" could be homemade; for example, one take required a balance that could be made from a coat hanger, plastic cups, and string. The Performance Assessment Administration Manual provided explicit instructions for setting up the equipment, described which tasks required servicing during administration, and contained instructions for recording information about the materials used that coders could refer to when scoring.

Students were required to move from station to station around a room to perform the tasks assigned to them. The administrator was responsible for overseeing the activities, keeping time, directing students to their stations, maintaining and replenishing equipment as necessary, and collecting the students' work. The administrator also provided advance instruction regarding certain materials and equipment, for tasks where the use of the equipment was not what was being measured. Administrators did not provide instruction on other procedures nor answer any other questions related to the activities required for the tasks.

To facilitate the students' movements around the room and keep track of where each should be, each student was given a routing card, prepared at the TIMSS national center. The routing cards stated the rotation scheme and sequence number of that student, his or her identifying information, and the stations to which the student was to go and in what order.

At each station, students performed the assigned task. This involved performing the designated activities, answering questions, and documenting their work in booklets (one booklet per task per student). Students had 30 minutes to work at each station. When students had finished their work at a station (or when time had expired), they handed their completed booklets to the administrator.

The performance assessment was not conducted in TIMSS-R.

Videotape Study. It was intended that TIMSS videotaping be spread out evenly over the school year. In Germany and the U.S. this goal was accomplished by employing a single videographer in each country to tape over an 8-month period, from October 1994 through May 1995. It was not possible to implement the same plan in Japan, due to the starting time of the school year in Japan and the necessity of coordinating the videotaping with the test administration. As a result, videotaping in Japan was compressed primarily into a 4-month period, from November 1994 though February 1995, with a few lessons taped in March.

Two kinds of data were collected in the TIMSS videotape study: videotapes and questionnaires. Supplementary materials deemed helpful for understanding the lesson (e.g., copies of textbook pages or worksheets) were also collected. Each classroom was videotaped once on a date convenient for the teacher. One complete lesson, as defined by the teacher, was videotaped in each classroom. Teachers were initially contacted by a project coordinator in each country who explained the goals of the study and scheduled the date and time for videotaping. Because teachers knew when the taping would take place, it was understood that they would attempt to prepare in some way for the event. In order to cut down somewhat on the variability in preparation methods across teachers, all participating teachers were given a common set of instructions, asking them not to make any special preparations for the taped class (e.g., by making special materials, planning special lessons, or practicing the lesson ahead of time). On the appointed day the videographer arrived at the school and videotaped the lesson. After the taping each teacher was given a questionnaire and an envelope in which to return it. The purpose of the questionnaire was to assess how typical the lesson was according to the teacher and to gather contextual information important for understanding the contents of the videotape.

All videotaping was done in real time, using a single camera. The camera was turned on at the beginning of the class, and not turned off until the lesson was over. In order to ensure comparability between videotapes, videographers were asked to adhere to two basic principles in choosing what to tape. The first principle required videographers to assume the perspective of an ideal student in the class and to aim the camera toward the object of focus of an ideal student at any given time.

An ideal student was defined as one who is always attentive to the lesson at hand and always occupied with the learning tasks assigned by the teacher, one who will attend to individual work when assigned to work alone, will attend to the teacher when she or he addresses the class, and will attend to peers when they ask questions or present their work or ideas to the whole class. In cases where different students in the same class are engaged in different activities, the ideal student is assumed to be doing whatever the majority of students are doing.

The second principle required videographers to capture everything the teacher did to instruct the class, regardless of the activities of the ideal student. Usually, this principle was in agreement with the first principle: whenever the ideal student is attending to the teacher, both principles would have the camera pointed at the teacher. However, there are times when the two principles are in conflict. In order to develop a set of standardized procedures for such instances, the three videographers were trained over the course of two intensive training seminars that lasted a total of 14 days. Tests conducted both during the training seminars and later during data collection revealed that videotaping methods were indeed comparable.

The TIMSS-R data collection methods differed in several respects from those used for TIMSS. Two cameras were used, instead of one, to videotape each lesson. One of the cameras focused primarily on the teacher, but was also used to capture close-ups of students' work during periods when students were working independently. The second camera was stationary. It was placed at the front of the room facing the students in order to capture students' interactions with the teacher and/or with each other during the lesson.

Editing. To maintain equality among countries, very little optical scanning and no image processing of item responses was permitted. All student test information was recorded in the student booklets or on separate coding sheets, and similar procedures were used for the questionnaires. Entry of the achievement and background data was facilitated by the *International Codebooks*, and the DataEntryManager software program.

The background questionnaires were stored with the various tracking forms so that the data entry staff could control the number of records to enter and transcribe the necessary information during data entry. NRCs were asked to arrange for double-entry of a random sample of at least 5 percent of the test instruments and questionnaires to gauge the error rate. An error rate of 1 percent was considered acceptable.

After entering data files in accordance with the international procedures, countries submitted their data files to the IEA Data Processing Center. There, TIMSS data underwent an exhaustive cleaning process designed to identify, document, and correct deviations for the international instruments, file structures, and coding schemes. The process also emphasized consistency of information with national data sets and appropriate linking among the many data files. The national centers were contacted regularly throughout the cleaning process and were given multiple opportunities to review the data for their countries. As a result of this review process, several items were identified as not being international comparable in certain countries and were deleted from the international data files and from the analyses for the international reports. In certain instances, recodes were performed on the cognitive items as a result of the item review.

Estimation Methods

Once TIMSS data are scored and compiled, the responses are weighted according to the sample design and population structure and then adjusted for nonresponse. This ensures that countries' representation in TIMSS is accurately assessed. The analyses of TIMSS data for most subjects are conducted in two phases: scaling and estimation. During the scaling phase, item response theory (IRT) procedures are used to estimate the measurement characteristics of each assessment question. During the estimation phase, the results of the scaling are used to produce estimates of student achievement (proficiency) in the various subject areas. The methodology of multiple imputations (plausible values) is then used to estimate characteristics of the proficiency distributions. Although imputation is conducted for the purpose of determining plausible values, no imputations are included in the TIMSS database.

Weighting. Appropriate estimation of population characteristics based on TIMSS samples requires that the TIMSS sample design be taken into account in all analyses. This is accomplished in part by assigning a weight to each respondent, where the sampling weight properly accounts for the sample design, takes into account any stratification or disproportional sampling of subgroups, and includes adjustments for nonresponse.

There are four types of sampling weights available for use with TIMSS data: student weights, school weights, student-teacher weights, and teacher weights. In all of these cases, weighted totals, means, and percentages using these weights are unbiased estimates of "weighted" national population totals, with the number of target grade students as the weight.

Student weights. The student sampling weights in TIMSS have two primary components: a student base weight and a nonresponse adjustment. The student base weight is the reciprocal of the student's probability of selection into the TIMSS sample, and is a product of up to three factors, reflecting the three stages of student sampling: the school selection probability, the classroom selection probability, and (if classroom subsampling has occurred) the student selection probability within selected classrooms. In most country samples, there is both school and student nonresponse. This nonresponse affects any estimators in that the effective sample size of both schools and students is reduced, increasing sampling variance. In addition, if there are systematic differences between the respondents and the nonrespondents, there will also be a bias of unknown size and direction in any estimators. This bias is partially adjusted for in TIMSS samples through the use of weighting adjustments multiplied to the student base weights.

Three versions of the students' sampling weight are provided in the user database. All three give the same figures for statistics such as means and proportions, but vary for statistics such as totals and population sizes. In addition to the total weight, described above, there are *House weights* and *Senate weights* for each student (the names are derived from an analogy with the U.S. legislative system). *House weights* are a set of weights based on the total sample size of each country, to be used when estimates across countries are computed or significance tests performed. The transformation of the weights will be different within each country, but in the end, the sum of the house-weight variables within each country will total to the sample size for that country. The house-weight variable is proportional to the total weight for that variable by the ratio of the sample size divided by the size of the population. These sampling weights can be used when the user wants the actual sample size to be used in performing significance tests.

Senate weights are a set of weights based on a constant scalar, to be used when estimates across countries are computed or significance tests performed. The transformation of the weights will be different within each country, but in the end, the sum of the senate-weight variables within each country will total to a fixed value (1000 in Populations 1 and 2, where two grades were sampled, and 500 in Population 3). The senate-weight variable, within each country, is proportional to the total weight by the ratio of 1000 (or 500) divided by the size of the population estimate. These sampling weights can be used when cross-national comparisons are required and the user wants to have each country contribute the same amount to the comparison, regardless of the size of the population.

Teacher weights. The teacher weight is a teacher-classroom weight, and so is greater than 0 for a classroom only if the teacher filled out a questionnaire for that classroom. The teacher-classroom weight is equal to the summation of the student-teacher weights for students linked to that classroom (for that assessment).

Student-teacher weights. The U.S. TIMSS public use file includes student-teacher weights and student-teacher replicate weights. These are aggregated into the teacher weights described above. Two student-teacher weights are assigned to each assessed student in U.S. TIMSS: a mathematics assessment weight and a science assessment weight. A student-teacher weight for a particular student and assessment is set to 0 if a teacher's questionnaire was not filled out for that student's assessment classroom. This occurred in the following situations: the teacher taught both mathematics and science and was randomly assigned to the other assessment; the teacher was assigned no classroom because of all his/her classrooms had fewer than five TIMSS-assessed students; the teacher was assigned a questionnaire classroom but not the student" classroom; the teacher refused to answer the questionnaire.

Population 3 advanced mathematics/physics adjustment factors. Student weights for Population 3 are similar to the Population 1 and 2 weights; but an additional set of weights was created to reflect the fact that some respondents had taken advanced mathematics or physics courses, or both. Weights were developed as the inverse of the probabilities that a student received a mathematics/physics literacy booklet, an advanced mathematics booklet, or a physics booklet. If a student was not assessed on these items, the value of the weight was set to 0. As a result, the total, house, and senate weights in Population 3 for each math or science assessment are the product of the base weight (the inverse of the school selection probability multiplied by the inverse of the student selection probability), the nonresponse adjustment factor, the literacy adjustment factor, the advanced mathematics adjustment factor, and the physics adjustment factor.

The internationally-defined weighting specifications for TIMSS-R require that each assessed student's sampling weight should be the product of (1) the inverse of the

BEST COPY AVAILABLE

school's probability of selection, (2) an adjustment for school-level nonresponse, (3) the inverse of the classroom's probability of selection, and (4) an adjustment for student-level nonresponse.

Scaling. The principal method by which student achievement is reported in TIMSS is through scale scores derived using IRT scaling. IRT is used to estimate students' average proficiency for the nation, for various subgroups of interest within the nation (e.g., those defined by age, race/ethnicity, sex), and for the states and territories. TIMSS utilized a one parameter IRT model to produce score scales that summarized the achievement results.

In 1999, the TIMSS-R assessment had five scales describing mathematics content strands and six scales for describing fields of science. The 1995 TIMSS data were rescaled using a three-parameter IRT model, to match the procedures used to scale the 1999 TIMSS-R data. After careful study of the rescaling process, the International Study Center concluded that the fit between the original TIMSS data and the rescaled TIMSS data met acceptable standards. However, as a result of rescaling, the average achievement scores of some nations changed from those initially reported in 1996.

Imputation. No imputations are generated for missing values in teacher, school, or questionnaires for any TIMSS data file. However, multiple imputation techniques have been applied to create *plausible values* for students' proficiency scores. The data include a set of five plausible values for each student in each of the assessed areas. Plausible values improve the estimation of population parameters at the cost of additional computational requirements.

Plausible values were developed during the analysis of the 1983–84 NAEP data in order to improve estimates of population distributions. In the TIMSS survey design, students are presented with separate blocks of exercises, each block consisting of both mathematics and science problems. Since each student attempts only a small portion of the total TIMSS test in each subject, attempts to estimate proficiency distributions are affected by the imprecision of the measurement. During the estimation phase, plausible values for content-area scale scores are generated for each student participating in the assessment. The plausible values technology estimates five possible scores for each student, which ensures that the estimates of the average performance of subpopulations and the estimates of variability in those estimates are more accurate and appropriate than if only a single score were estimated for each student.

The process of drawing plausible values from the predictive distribution of proficiency values is called "conditioning." Plausible values are computed separately for each population. They are based on the student's responses to the items going into the scale and on the values of a set of background variables that are important for the reporting of proficiency scores. The variables used to calculate plausible values for a given assessment scale or group of scales include a broad spectrum of background, attitude, and experiential variables and composites of such variables.

Rubin (1987) proposes that this process be carried out several times—that is, multiple imputations—so that the uncertainty associated with imputation can be quantified.

Future Plans

Another international assessment—Trends in International Mathematics and Science Study—is currently planned for 2003, and will survey both 4th- and 8th-grade students. Subsequent follow ups are planned at 4-year intervals thereafter.

5. DATA QUALITY AND COMPARABILITY

In addition to setting high standards for data quality, the TIMSS International Study Center has tried to ensure the overall quality of the study through a dual strategy of support to the national centers and quality control checks.

Despite the efforts taken to minimize error, any sample survey as complex as TIMSS has the possibility of error. Below are discussed possible sources of error in TIMSS.

Sampling Error

With complex sampling designs that involve more than simple random sampling, as in the case of TIMSS where a multistage cluster design was used, there are several methods for estimating the sampling error of a statistic that avoid the assumption of simple random sampling. One such method is the jackknife repeated replication (JRR) technique. The particular application of the JRR technique used in TIMSS is termed a paired selection model because it assumes that the sampled population can be partitioned into strata, with the sampling in each stratum consisting of two primary sampling units (PSUs) selected independently. Following this first-stage sampling, there may be any number of subsequent stages of selection that may involve equal or unequal probability selection of the corresponding elements. The TIMSS design called for a total of 150 schools for the target population. These schools constituted the PSUs in most countries, and were paired sequentially after sorting by a set of implicit stratification variables. This resulted in the implicit creation of 75 strata, with two schools selected per stratum.

Imputation error. The variance introduced by imputation of missing data must be considered when using plausible values to estimate standard errors for proficiency estimates. The general procedure for estimating the imputation variance using plausible values is as follows: first estimate the statistic (t), each time using a different set of the plausible values (M). The statistics t_m can be anything estimable from the data, such as a mean, the difference between means, percentiles, etc. If all of the (M=5) plausible values in the TIMSS database are used, the parameter will be estimated five times, once using each set of plausible values. Each of these estimates will be called t_m, where m=1,2,...,5. Once the statistics are computed the imputation variance is then computed as:

$$Var_{imp} = \left(1 + \frac{1}{M}\right) \cdot Var(t_m)$$

where M is the number of plausible values used in the calculation and $Var(t_m)$ is the variance of the estimates computed using each plausible value.

Nonsampling Error

Due to the particular situations of individual TIMSS countries, sampling and coverage practices had to be adaptable, in order to ensure an internationally comparable population. As a result, nonsampling errors in TIMSS can be related to both coverage error and nonresponse. Measurement error was also a nontrivial issue in administering TIMSS, as different countries had different mathematics and science curricula. These potential sources of error are discussed in detail below.

Coverage error. The stated objective in TIMSS was that the effective population, the population actually sampled by TIMSS, be as close as possible to the International Desired Population. Yet, because a purpose of TIMSS was to study the effects of different international curricula and pedagogical methods on mathematics and science learning, participating countries had to operationally define their population for sampling purposes. Some NRCs had to restrict coverage at the country level, for example, by excluding remote regions or a segment of the educational system. In these few situations, countries were permitted to define a national desired population that did not include part of the International Desired Population. Exclusions could be based on geographic areas or language groups. Most countries participating in the Population 3 (20 out of 24) had 100 percent coverage, after sample exclusions. Among the four countries with incomplete coverage, the coverage rate ranged from 50 percent for Latvia to 84 percent for Lithuania.

To provide a better curricular match, several Population 2 countries elected to test students in the 7th and 8th grades (the two grades tested by most countries), even though that meant not testing the two grades with the most age-eligible students. This led to the students in these four countries being somewhat older than those in the other countries. The majority of countries in all sample populations satisfied the international guidelines for sample participation rates, grade selection, and sampling procedures.

Nonresponse error.

Unit nonresponse. Unit nonresponse error results from nonparticipation of schools and students. Weighted and unweighted response rates were computed for each participating country by grade, at the school level, and at the student level. Overall response rates (combined school and student response rates) were also computed.

The minimum acceptable school-level response rate, before the use of replacement schools, was set at 85 percent. This criterion was applied to the unweighted school-level response rate. Both weighted and unweighted school-level response rates were reported, with and without replacement schools. It was generally the case that weighted and unweighted response rates were similar.

Like the school-level response rate, the minimum acceptable student-level response rate was set at 85 percent. This criterion was applied to the unweighted student-level response rate. Both weighted and unweighted student level response rates were calculated. The weighted student-level response rate is the sum of the inverse of the selection probabilities for all participating students divided by the sum of the inverse of the selection probabilities for all eligible students.

Measurement error. Measurement error is introduced into a survey when its test instruments do not accurately measure the knowledge or aptitude they are intended to assess. The largest potential source of measurement error in TIMSS results from differences in the mathematics and science curricula across participating countries. In order to minimize the effects of measurement error, TIMSS carried out a special test called the Test-Curriculum Matching Analysis (TCMA). Each country was asked to identify, for each item, whether the topic of the item was intended in the curriculum for the majority of the students.

Data Comparability

The data collected for TIMSS in 1995 and the data collected for TIMSS-R in 1999 are comparable because comparability was built into the design and implementation. Through a careful process of review, analysis, and refinement, the assessment and questionnaire items were purposefully developed and field tested for similarity and for reliable comparisons between TIMSS and TIMSS-R. After careful review of all available data, including a test for item reliability between old and new items, the TIMSS and TIMSS-R assessments were found to be very similar in format, content, and difficulty level. Moreover, TIMSS and TIMSS-R data are on the same 8th-grade scale to allow for reliable comparisons between the two 8th-grade cohorts over time. Procedures for conducting the assessments were the same.

Findings from comparisons between the results of TIMSS and TIMSS-R, however, cannot be interpreted to indicate the success or failure of mathematics and science reform efforts within a particular country, such as the United States. TIMSS-R was designed to specifications detailed in the TIMSS curriculum frameworks. International experts developed the TIMSS curriculum frameworks to portray the structure of the intended school mathematics and science curricula from many nations, not specifically the United States Thus, when interpreting the findings, it is important to take into account the mathematics and science curricula likely encountered by U.S. students in school. TIMSS and TIMSS-R results are most useful when they are considered in light of other knowledge about education systems, including not only curricula, but also factors such as trends in education reform, changes in the school-age populations, and societal demands and expectations.

The ability to compare data across different countries constitutes a considerable part of the purpose behind TIMSS. As a result, it was crucial to ensure that items developed for use in one country were functionally identical to those used in other countries. Because questionnaires were originally developed in English and later translated into the language of each of the TIMSS countries, some differences do exist in the wording of questions. NRCs from each country reviewed the national adaptations of individual questionnaire items and submitted a report to the IEA Data Processing Center. In addition to the translation verification steps used for all TIMSS test items, a thorough item review process was used to further evaluate any items that were functioning differently in different countries according to the international item statistics. In certain cases, items had to be recoded or deleted entirely from the international database as a result of this review process.

6. CONTACT INFORMATION

For content information about TIMSS, contact:

Patrick Gonzales
Phone: (202) 502–7346
E-mail: patrick.gonzales@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

Most of the technical documentation for TIMSS is published by Boston College. The U.S. Department of Education, National Center for Education Statistics, is the source of several additional references listed below; these publications are indicated with an NCES number.

General

Pursuing Excellence: Comparisons of International Eighth-Grade Mathematics and Science Achievement from a U.S. Perspective, 1995 and 1999, NCES 2001–028, by P. Gonzales, C. Calsyn, L. Jocelyn, K. Mak, D. Kastberg, S. Arafeh, T. Williams, and W. Tsen. Washington, DC: 2000.

Uses of Data

Linking the National Assessment of Educational Progress (NAEP) and The Third International Mathematics and Science Study (TIMSS): A Technical Report, NCES 98–499, by E.G. Johnson. Washington, DC: 1998.

Linking The National Assessment of Educational Progress (NAEP) and The Third International Mathematics and Science Study (TIMSS): Eighth-Grade Results, NCES 98–500, by E.G. Johnson and A. Siegendorf. Washington, DC: 1998.

User's Guide for the Third International Mathematics and Science Study (TIMSS) and U.S. Augmented Data Files, by B. Chaney, L. Jocelyn, D. Levine, T. Mule, L. Rizzo, K. Rust, S. Roey, T. Williams, and S. Warren. Rockville, MD: 1998.

Survey Design

Multiple Imputation for Nonresponse in Surveys, by D.B. Rubin. New York: John Wiley & Sons, 1987.

TIMSS International Study Center, Boston College, *TIMSS Technical Report: Volume I: Design and Development*, by M.O. Martin and D.L. Kelly (eds.). Chestnut Hill, MA: 1996.

TIMSS International Study Center, Boston College, *TIMSS Technical Report: Volume II: Implementation and Analysis Primary and Middle School Years*, by M.O. Martin and D.L. Kelly (eds.). Chestnut Hill, MA: 1998.

TIMSS International Study Center, Boston College, *TIMSS Technical Report: Volume III: Final Year of Secondary School*, by M.O. Martin and D.L. Kelly (eds.). Chestnut Hill, MA: 1998.

Data Quality and Comparability

TIMSS International Study Center, Boston College, *Quality Assurance in Data Collection*, by M.O. Martin and I.V.S. Mullis (eds.). Chestnut Hill, MA: 1996.

Chapter 22: IEA Reading Literacy Study

1. OVERVIEW

STUDY OF 9-YEAR-OLD AND 14-YEAR-OLD STUDENTS IN 32 COUNTRIES

IEA Reading data collected through:

- Reading Assessment
- Student Questionnaire
- Teacher Questionnaire
- Principal Questionnaire
- National Questionnaire

The International Association for the Evaluation of Educational Achievement (IEA) Reading Literacy Study was conducted during the 1990–91 school year in 32 countries around the world. The International Steering Committee (ISC), the International Coordinating Center (ICC), and the National Research Coordinators of each of the participating countries developed the assessment instruments, assessment procedures, and scaled scores used to report the results and oversaw the conduct of the study internationally. Nationally representative samples of the classes in the grades with the most 9-year-old and 14-year-old students were directed to read and respond to a broad range of materials over two testing periods. The U.S. component involved 7,200 4th-grade students and 3,800 9th-grade students at 332 public and private schools, distributed in 227 districts across 31 states and the District of Columbia.

Purpose

To (1) develop internationally valid instruments for measuring reading literacy suitable for establishing internationally comparable literacy levels in each of the participating countries; (2) describe on one international scale the literacy profiles of 9- and 14-year-olds in school in each of the participating countries; (3) describe the reading habits of the 9- and 14-year-olds in each participating country; and (4) identify the home, school, and societal factors associated with the literacy levels and reading habits of the 9-year-olds in school.

Components

The IEA Reading Literacy Study used a reading assessment instrument and four sets of questionnaires (for students, their teachers, their principals, and the nation) developed by committees working under the International Sampling Coordinator. The instruments were designed so that the same content would be used in all participating countries in the appropriate languages for those countries.

Reading Literacy Tests. Two reading assessments were developed to measure the reading proficiency of 9- and 14-year-olds. The assessments were designed to provide scaled scores that reflect students' understanding of three types of text: narrative prose (continuous text materials in which the writer's aim was to tell a story, whether fact or fiction), expository prose (continuous text materials designed to describe or explain things), and documents (structured tabular texts, such as forms, charts, labels, graphs, lists, and sets of instructions). The assessments include questions that tapped six types of reading processes: verbatim, paraphrase, inference, main theme, locating information, and following directions.

Questionnaires. The four sets of questionnaires—student, teacher, principal, and national—were designed to collect data about those factors that are known to influence reading achievement and that might vary across nations. These data could best be described in terms of two dimensions: to whom and to what they referred. In the case of the who dimension, the data describe students, their families, their teachers, and their schools. On the what dimension, the data describe their attributes, the kinds of environments provided, the forms of instruction used, and the reading behaviors they exhibited.

Student Questionnaires included items on student/parent background information such as parent's educational level, language spoken at home, student reading activities, etc. There were separate questionnaires for 4th and 9th graders.

Teacher Questionnaires were used to collect information on school and classroom policy, instructional approaches used by the teacher, and the teacher's educational background and experience.

School Questionnaires were completed by the school principal or person designated by the school principal on school demographics, school policies and resources, and evaluation of instruction. One questionnaire was to be obtained from each participating school.

The *National Questionnaire*, completed by the national research team, was used to collect data about the national system, and requested data on standard demographic characteristics, available resources, and practices related to reading achievement.

Periodicity

The IEA Reading Literacy Study was conducted in 1991. The Progress in Reading Literacy Study (PIRLS) was administered in 2001 and tested just 4th-grade students.

2. USES OF DATA

Beyond the usual reporting of reading literacy in NCES compendia (e.g., *Digest of Education Statistics*, *Youth Indicators)*, NCES released four volumes concerning the IEA Reading Literacy Study. These include a technical report, a methodological report, a summary of findings, and a set of collected papers. Among the issues discussed in these reports are sampling for international comparative studies in education, the development and interpretation of reading literacy scales, the study of various effects (e.g., classroom, school, community, family) on reading literacy, and instructional practice in teaching reading.

3. KEY CONCEPTS

Some of the key concepts related to the IEA Reading Literacy Study are described below.

Types of text. Scaled scores were developed to reflect students' understanding of three types of text:

Narrative prose. Continuous text materials in which the writer's aim was to tell a story, whether fact or fiction. They are normally designed to entertain or involve the reader emotionally; they are written in the past tense, and usually have people or animals as their main theme;

Expository prose. Continuous text materials designed to describe or explain something. The subjects of such text are usually things, but they may be written in the present or the past; the style is typically impersonal, highlighting such features as definitions, causes, classifications, functions, contrasts, and examples, rather than a moving plot with climax; and

Documents. Structured tabular texts, such as forms, charts, labels, graphs, lists, and sets of instructions where the reading requirements typically involve locating information or following directions, rather than continuous reading of connected text.

4. SURVEY DESIGN

Target Population

Within each of the participating countries, nationally representative samples were to be drawn based on two internationally defined target populations: (1) *Population A*: All students attending school on a full-time basis at the grade level in which most students 9 years old (during the 1st week of the 8th month of the school year) are enrolled; and (2) *Population B*: All students attending school on a full-time basis at the grade level in which most students 14 years old (during the 1st week of the 8th month of the school year) are enrolled.

Within the United States, these definitions were implemented and modified in the following ways: (1) *Population A*: All students attending school on a full-time basis at the grade 4 level in the 50 states and the District of Columbia, during the 1990–91 school year, who, in the opinion of school personnel, are capable of taking the test; and (2) *Population B*: All students attending school on a full-time basis at the grade 9 level in the 50 states and the District of Columbia, during the 1990–91 school year, who, in the opinion of school personnel, are capable of taking the test.

BEST COPY AVAILABLE

A number of practical sampling issues in the United States necessitated some additional departures from the procedures proposed in the IEA sampling manual (Ross 1991). First, because the geographic dispersion of schools made it fiscally impossible to consider collecting data from a stratified random sample of schools, the sample size was increased to offset the additional clustering effects introduced by the three-stage sampling frame designed to facilitate data collection. Second, because the United States lacks a single set of national policies that would control such factors as entrance age, retention in grades, and placement in mainstream classes, study designers in the United States could not identify a single grade with a clean majority of the target population. Hence, the national target population was defined so that the modal grade for each desired age group was chosen. These modal grades contained more than 50 percent (i.e., a majority) of students of the relevant age in each case.

Sample Design

The sample for the IEA Reading Literacy Study was selected using a complex multistage clustered design involving the sampling of intact classes from selected schools within selected geographic areas, called primary sampling units (PSUs), across the United States.

The structure of the sampling design differed somewhat from the models suggested by the international referee (Ross 1991). The United States adopted the approach, approved by the referee, of arranging for personnel from outside the school system to administer the assessments. This approach was taken to maximize school participation by minimizing the burden on schools and to assist in maintaining uniformly high standards of assessment administration throughout the sample by using field workers who were trained as a group by study staff. In most other countries, school personnel administered the assessments in the interest of minimizing costs.

The basic U.S. sample plan called for sampling intact classrooms and/or classes. For grade 4, if a sample school had fewer than an estimated 50 4th-grade students, all were included. In schools with 50 or more 4th graders, two classrooms were taken at random. For grade 9, in schools with fewer than an estimated 25 9th-grade students, all were included. Otherwise, the plan called for taking one classroom (typically, the language arts class). The number of students in the grade was estimated by dividing the total enrollment, as reported on the 1989 Quality Education Data (QED) file, by the grade span of the school.

The multistage sampling process for the IEA Reading Literacy Study involved the following steps:

(1) Selection of PSUs

(2) Selection of schools (public and nonpublic) within the selected PSUs

(3) Selection of intact classrooms and/or classes within the selected schools

Selection of PSUs. In the first stage of sampling, the United States (the 50 states and the District of Columbia) were divided into the geographic PSUs used by the National Assessment of Educational Progress (NAEP), which are counties (or independent cities) and groups of counties with a minimum population of 60,000 as of the 1980 Census. The counties composing metropolitan areas are kept together; other aggregations avoid mixing urban and rural counties. Since IEA specifications did not require certain estimates by subgroups (such as minorities) that were mandated by NAEP, the NAEP PSUs were restratified for use in the IEA study. The first level stratification was by NAEP region (four geographic strata) and two degrees of urbanization strata (Metropolitan Statistical Area—MSA—and non-MSA). In addition, the Southeast and West regions were stratified by percent minority, those with less than 20 percent minorities in one class and those with 20 percent or more in another.

Fourteen PSUs were of sufficiently large size that it was appropriate to include them in the sample with certainty. Minorities (outside of the large cities, included with certainty) are relatively less prevalent in the Northeast and the Central regions, so the minority stratification was not used in those regions. The high minority, non-MSA stratum in the West contained so few PSUs that it was combined with the low minority, non-MSA stratum. It was possible to subdivide them by percent minority in the second stage of stratification.

A sample of 50 PSUs in total was drawn according to the above allocation. Sampling weights equal to the inverse of the probabilities of selection were attached to them.

Selection of schools. The schools in the sampled PSUs were extracted from the QED file and were substratified by stage II strata. The two stage II stratifying variables were type of control (public schools in one class; private schools in the other class) and enrollment in the 4th grade for Population A or the 9th grade for Population B.

The schools were put into three classes at Population A and two classes at Population B on the basis of their estimated grade enrollment. A relatively thin sample of small schools was drawn to increase the efficiency of the design, since the per-student assessment costs for such schools were high. This had the effect of increasing the weights of these schools so that their effect on national projections was proportionate to the total enrollment of the stratum.

The sample of 200 schools from each population was allocated to the deeply stratified universe in proportion to the number of students in the given grade projected from the sampled PSUs, since, at the time the sample was drawn, total counts for the universe were not available in time to meet the deadline for the design work. This required a later adjustment in the sampling weights, as is discussed later in this section.

As required by the sampling referee, checks were made on the selected sample of schools and their base weights to ensure that the samples had been drawn without error. By stratum, the weighted measures of size of the selected schools were summed and then compared with the total of the measures of size for the stratum. They agreed exactly in each case, as was appropriate.

Selection of intact classroom and/or classes. As schools agreed to participate in the IEA study, they were sent a Fourth/Ninth Grade Class List Form asking for names and identifying information for all eligible classes within that school. This Class List Form was used to select the sample of the class(es) participating in the study.

Data Collection and Processing

The National Center for Education Statistics (NCES) began its efforts to gain support for the IEA Reading Literacy Study through presentations to the Council of Chief State School Officers' (CCSSO) Education Information Advisory Council (EIAC). EIAC endorsed the study and encouraged its members to participate fully in all activities.

According to the specifications of IEA, those who would conduct the Reading Literacy Study should first obtain permission to test in the schools. In the United States, because the school system is decentralized and locally autonomous, this requirement necessitated adherence to a protocol of contacting several levels of government officials: chief state school officers, local district superintendents, building principals, and classroom teachers.

The IEA Reading Literacy Study was administered by Westat, under a contract administered by NCES. Westat selected the schools in the sample and made the necessary contacts with state, district, and school administrators to obtain permissions to test in these schools. It also recruited, trained, and supervised the field assessment staff, and received the completed materials.

Reference dates. Data for the IEA Reading Literacy Study were collected in February and March, 1991.

Data collection. The ICC specifications permitted participating countries to choose field administrators from a range of categories, including classroom teachers, school administrators, and nonschool personnel. The U.S. study team felt that the study would be better served by creating a field staff that was in no way associated with the schools themselves. The primary benefit would be that the assessment administrators could be trained together and would subsequently administer the test to all students in a standardized manner. In addition, using study staff rather than school personnel would reduce the burden of response and might thereby increase the rate of participation.

Subsequently, Westat hired and trained a field staff of 45 assessment administrators and two supervisors to administer and collect the data. Each assessment administrator met with a coordinator at each school to schedule the assessments and make appropriate arrangements. At this time, it was also determined which students appearing on the class roster should be identified as "excluded" on the Administration Schedule. For this study, a student was excluded from the assessment only for the following two reasons: (1) a student was enrolled in a special education program and had an Individual Educational Plan (IEP) that specifically prohibited pencil-and-paper assessment; or (2) a student was non-English speaking and had been enrolled in a mainstream English class for less than 2 years. In total, 183 students were excluded from the grade 4 sample and 18 students from the grade 9 sample.

Each set of classroom sessions involved approximately 25 students, each of whom completed the Reading Literacy Test and the Student Questionnaire.

Data was collected on approximately 7,200 students in the 4th grade and 3,800 students in the 9th grade, with 167 schools participating at grade 4 and 165 at grade 9. Both public and private schools were included, distributed in 227 districts across 31 states and the District of Columbia. Three hundred 4th-grade and 160 9th-grade teachers also provided data for the study, as did 332 school administrators.

Data processing. Those materials returned directly to Westat included the School Questionnaire, the Teacher

Questionnaire, and the Student Questionnaires. The assessment administrators sent the Reading Literacy Tests to Data Recognition Corporation (DRC) for coding, keying, verifying, and basic editing.

The data keying at Westat used a 100 percent verification system. All data were entered twice by different operators and then compared. Any differences were resolved, with the supervisor adjudicating difficult cases. After keying, additional machine editing was used to detect and resolve range and logic errors.

DRC had two major tasks: key entry of the responses to the Reading Literacy Test items (DRC also used a 100 percent verification system), and scoring the open-ended writing responses included in the Reading Literacy Tests. Each essay was read by two readers independently and scored; if the scores differed, a third resolving reading was done by a task leader. Scoring was monitored closely, with daily reports produced for each reader indicating the number of papers read, the percentage of exact, adjacent, and nonadjacent agreement with the other readers of the same papers, the tendency of the disagreement, and the score point distribution. The area of scrutiny was inconsistency, or drift from an established standard. Throughout the project, readers scored sample papers at rangefinding meetings in order to validate and recalibrate the criteria. Retraining was ongoing to secure continued familiarity with and adherence to the scoring criteria and to prevent roomwide drift as the project progressed. Legibility issues were addressed implicitly in the open-ended question scoring process.

The scorers of the open-ended items were experienced in scoring similar questions for other large-scale assessments. They were generally high school teachers who were provided training for scoring open-ended questions for this study.

Editing. The first phase of data editing took place during the keying of the questionnaires and literacy assessments. The 100 percent verification process required all data to be entered twice by different operators and then compared. Discrepancies were corrected, and in the case of difficult cases, were adjudicated by the supervisor.

In the second phase of data editing, a machine-edit program was used to detect and resolve as many errors as possible prior to delivering the data for more complex interfile editing and statistical data quality analyses. The errors detected by machine editing were of two general types: (1) range errors, in which response values fell outside a predetermined acceptable range; and (2) logic errors, in which there were some inconsistencies between response values. These included improperly followed skip patterns, data inconsistencies among two or more variables, and addition checks where values of a group of variables were to sum to a known value.

Creating the files. The study produced eight U.S. files in all. Two were reading test data for each population. In addition, a file was created for each population for the Student, Teacher, and School Questionnaires.

These eight U.S. files were combined and reformatted in accordance with the specifications provided by ICC to produce six ICC international format files. The U.S. Teacher and School Questionnaire files were mapped onto ICC versions; the U.S. Student Questionnaire and Reading Literacy Test files were mapped onto a single ICC student file for each population. While only a few of the questions in the U.S. questionnaires were asked with the same wording and response alternatives as their analogues in the ICC version, the data, nonetheless, were to go to the ICC in the format of its questionnaires.

The ICC supported its questionnaires with software for data entry, record editing, range checks, ID checks across files, and logic and consistency checks, including skip patterns and intra- and interfile checks. When the data were converted to ICC format and these checking programs were run, almost all of the errors occurred in cases where a prescribed range was violated by a legitimate, if unusual, value, or a consistency check was violated by a combination of such values. Essentially the data did not require further editing in order to conform to ICC standards.

As part of the agreement to participate in IEA Reading Literacy Study, each participating country, including the United States, had granted IEA permission to release its data to individuals or organizations desiring to perform secondary analyses. To avoid disclosure problems, the U.S. files submitted to IEA were considered public use data files, and extensive analyses were performed to ensure that individual respondents could not be identified.

Estimation Methods

Once IEA data were scored and compiled, the responses were weighted according to the sample design and population structure and then adjusted for nonresponse. This ensured that the students' representation in the IEA Reading Literacy Study matched their actual proportion in the school population for the grades assessed.

Weighting. The weighting of the national IEA sample reflected the probability of selection for each student in the sample, adjusted for nonresponse. The weight assigned to a student's responses was the inverse of the probability that the student would be selected for the sample. Through poststratification, weighting ensured that the representation of certain subpopulations corresponded to figures from the Current Population Survey (CPS) and also accounted for the low sampling rates that occurred for very small schools. Thus, properly weighted IEA data provided results that reflect the representative performances of the entire nation and of the subpopulations of interest. The following provides an overview of the steps involved in deriving the sampling weights.

Applying the secondary stratification only to the schools in the initial sample of NAEP PSUs, after weighting the characteristics of the schools in the sampled PSUs by the inverse of the probabilities of selection of those PSUs, introduced sampling error in the estimates of the substratum totals. Since the time that the design was set, it has been possible to tabulate the entire QED file by the characteristics that define the substrata. This made it possible to adjust the sample weights so that the number of schools in the selected sample would weight up to the number of schools in the QED tape within each substratum—a straightforward poststratification procedure.

The enrollments in the sampled schools were multiplied by the school weights and compared with estimated enrollments for the 4th and 9th grades produced by the CPS. The differences were judged to be large enough that a second adjustment to the sampling weights was made so that the estimated enrollments in the two grades would equal the CPS estimates within each NAEP region.

The two weight adjustments automatically corrected for school nonresponse to the survey. In making the first adjustment, the weighted number of sampled schools was adjusted to equal the number of schools listed in the QED file, with no account taken of the number of schools that had closed.

The student weights within each school reflected both the subsampling of classrooms in the school and the individual student nonresponse within the school. That is, the school weight was multiplied by the number of classrooms in the school and divided by the number of classrooms sampled. This weight was multiplied by the number of students in the selected classrooms and divided by the number of responding students to produce the student weights.

Scaling. For purposes of summarizing item responses, the ISC developed procedures for creating international scaled scores based on the Rasch model, the one-parameter item response theory (IRT) model. The underlying principle of IRT is that, when a number of items require similar skills, the regularities observed across patterns of response can often be used to characterize both respondents and tasks in terms of a relatively small number of variables.

The ICC performed all tasks related to scaling of the Reading Literacy Tests (i.e., calibrated items and estimated student abilities). Calibration of items and estimation of abilities were performed separately for each of the three reading literacy domains (narrative, expository, and document). Item difficulties were estimated on the basis of responses of a random sample of students selected from all participating countries. This international calibration sample consisted of 10,790 students for grade 4 and 10,772 for grade 9.

The ICC deleted a total of six items for grade 4 and seven items for grade 9 that did not fit the international calibration sample. Rasch analysis was performed within each participating country, setting the item difficulties derived on the international calibration sample as known parameters. Item fit was also examined within each participating country. If an item was found not to fit the Rasch model in a given country, that item was not included in estimating student abilities within the country under consideration. Based on the invariance properties of the Rasch model (i.e., examinee ability estimation is independent of the particular set of items administered from a calibrated pool), the ICC derived reading literacy ability estimates for students within each participating country and placed them on a common scale. For ease of use, the logit scale was transformed such that the international mean and standard deviation were 500 and 100, respectively, for each reading literacy domain.

Since the international mean and standard deviation were arbitrarily set, the scale scores across the domains are not equated. Similarly, the scale scores across the two populations are not equated either.

Imputation. The IEA study employed a combination of a hot-deck imputation procedures and deterministic imputations to assign values for missing responses for the data items. Hot-deck (using Wesdeck) imputation procedures were used to handle missing responses for most items. For some of the remaining items, the missing responses were completed from information available in other data sources; for some items, it was

BEST COPY AVAILABLE

possible to deduce the missing response from the responses to other items on the questionnaire; and for other items, the overall modal response for respondents was assigned for all missing responses. The latter technique, which was employed for operational expediency, was used only when the item nonresponse rate was very small.

Future Plans

The IEA plans to continue its study of reading literacy through PIRLS, an assessment of 4th graders on a recurring basis.

5. DATA QUALITY AND COMPARABILITY

The U.S. component of the IEA Reading Literacy Study had to report accurate results for populations of students and subgroups of these populations (e.g., minority students or students attending nonpublic schools). Although only a very small percentage of the student population in each grade were assessed, IEA Reading Literacy Study estimates are accurate because they depend on the absolute number of students participating, not on the relative proportion of students.

Every activity in IEA Reading Literacy Study assessments was conducted with rigorous quality control. All questions underwent extensive reviews by subject-area and measurement specialists, as well as careful scrutiny to eliminate any potential bias or lack of sensitivity to particular groups. The complex process by which IEA Reading Literacy Study data were collected and processed was monitored closely. Westat ensured uniformity of procedures through training, supervision, and quality control monitoring. (See section 4 for more detail on quality control procedures.)

With any survey, however, there is the possibility of error. The most likely sources of error in the IEA Reading Literacy Study are described below.

Sampling Error

The primary component of uncertainty in the IEA Reading Literacy Study is due to sampling only a small number of students relative to the whole population. This accounts for the variability of estimates of percentages of students having certain background characteristics or answering a certain cognitive question correctly.

Because the IEA Reading Literacy Study used complex sampling procedures, a jackknife replication procedure was used to estimate standard errors. A set of jackknife replicate weights was developed for each assessed student.

Because of the effects of clustering and unequal probabilities of selection in the IEA Reading Literacy Study, in most cases the design effect is greater than 1. This means that the sample design is generally less efficient than simple random sampling, although it is more cost-effective.

Nonsampling Error

While there is the possibility of some coverage error in the IEA Reading Literacy Study, the two most likely types of nonsampling error are nonresponse error due to nonparticipation and measurement error due to instrumentation defects (described below). The overall extent of nonsampling error is largely unknown.

Coverage error. In the IEA Reading Literacy Study, coverage error could result from either the sampling frame of schools being incomplete or from the schools' failure to include all the students on the lists from which grade samples were drawn. The IEA Reading Literacy Study, while conducted in 1991, used the 1989 QED school list for the names of the regular public and private schools. This list, however, did not include schools that opened between 1989 and the time of the 1991 IEA Reading Literacy Study. The weighting adjustment for school nonresponse to the survey considered schools closed between 1989 and 1991 as nonresponding schools. Apparently there was no check by the assessment administrators to verify the inclusion of all students on the lists provided them.

Nonresponse error. Unit nonresponse error results from nonparticipation of schools and students. Item nonresponse error results from students who participate but do not answer every question.

Unit nonresponse. The unweighted school response rate across public and private sectors was 87 percent for the grade 4 schools and 86 percent for the grade 9 schools. These rates exceeded the international requirement of at least 85 percent for each grade. At the student level, about 7 percent of the grade 4 students and 14 percent of the grade 9 students were unit nonrespondents. Weighting class adjustments were used to compensate for unit nonresponse at both the school and student levels. There were responses from all teachers and administrators (100 percent response rate) on the teacher and administrator questionnaires, so no adjustments were necessary to compensate for unit nonresponse on these two sets of data.

BEST COPY AVAILABLE

Item nonresponse. Item nonresponse to the questionnaire items occurred when a student who completed the reading performance test failed to complete an item on the student background questionnaire, or when a teacher or principal failed to complete an item on the questionnaires that they completed. The level of item nonresponse was generally low, but some items were not answered by 10 percent or more of the respondents.

Data Comparability

Since the IEA Reading Literacy Study was by definition an international study involving 32 countries, it allows comparisons between participating countries. Additionally, the results of the IEA Reading Literacy Study should be comparable with those of the NAEP Reading assessments. Trend comparisons are available through PIRLS.

Comparisons with other countries. In contrast to the poor showing of American students in other international comparisons, in reading, at least, American students were among the best of the 32 nations involved in the study. With the exception of Finland, no country consistently outperformed the United States. It should be noted that these 32 nations are a self-selected group that are neither a representative sample of all nations nor of our principal trading partners (e.g., Japan, the United Kingdom, and Mexico were not included). However, among these are 18 members of the Organization for Economic Cooperation and Development (OECD), and the average of the OECD countries is a benchmark against which measurements of the overall American performance, as well as particular American subpopulations, can be compared. This has been done in the NCES report *Reading Literacy in the United States: Findings from the IEA Reading Literacy Study* (NCES 96-258). The NCES report *Reading Literacy in an International Perspective: Collected Papers from the IEA Reading Literacy Study* (NCES 97-875) contains nine papers addressing issues regarding reading literacy, focusing on outcomes in literacy achievement, instructional practices in reading, and school climate. Several of these papers limit their analysis to a nine-country focus of eight European nations and the United States.

Comparisons with NAEP Reading assessments. The finding that the results of the IEA study were more optimistic in their portrayal of the reading proficiency of American students than the results of the NAEP assessments has generated additional study comparing the two assessments in an effort to determine the reason for these differences. (See chapter 20.)

Comparisons with PIRLS. The PIRLS data collection was scheduled for 2001 to coincide with the 10th anniversary of the IEA Reading Literacy Study to provide an opportunity for countries that participated in the earlier study to obtain a measure of change from 1991. The United States was among the countries that participated in the PIRLS trend study, in which the 1991 test and student questionnaire were administered to a sample of PIRLS students.

Content changes. For PIRLS in 2001, the general thrust of the assessment was the same, although the frameworks were modified and new test items were developed.

Design changes. Given that a large number of countries which are participating in PIRLS are also participating in the OECD Program for International Assessment (PISA), the older cohort has been eliminated. Only one age/grade level is being tested.

6. CONTACT INFORMATION

For content information on the IEA Reading Literacy Study, contact:

Eugene Owen
Phone: (202) 502-7422
E-mail: eugene.owen@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006-5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Reading Literacy in the United States: Technical Report, NCES 94-259, by M. Binkley and K. Rust (eds). Washington, DC: 1994.

Survey Design

Sampling Manual for the IEA International Study of Reading Literacy, by K.N. Ross. University of Hamburg, Hamburg, Germany: International Coordinating Center, IEA International Study of Reading Literacy, 1991.

Data Quality and Comparability

Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study, NCES 94-469, by M. Binkley, K. Rust, and M. Winglee (eds). Washington, DC: 1995.

Chapter 23: National Adult Literacy Survey (NALS)

1. OVERVIEW

The National Adult Literacy Survey (NALS) was initiated to fill the need for accurate and detailed information on the English literacy skills of America's adults. In accordance with a congressional mandate, it provides the most detailed portrait that has ever been available on the condition of literacy in this nation—and on the unrealized potential of its citizens.

The 1992 National Adult Literacy Survey is the third and largest assessment of adult literacy funded by the federal government and conducted by the Educational Testing Service (ETS). The two previous efforts were: (1) the 1985 Young Adult Literacy Assessment (funded as an adjunct to the National Assessment of Educational Progress—see chapter 20); and (2) the Department of Labor's 1990 Workplace Literacy Survey. Building on these two earlier surveys, literacy for the NALS is defined along three dimensions—prose, document, and quantitative—designed to capture an ordered set of information-processing skills and strategies that adults use to accomplish a diverse range of literacy tasks encountered in everyday life. The background data collected in NALS provide a context for understanding the ways in which various characteristics are associated with demonstrated literacy skills.

NALS is the first national study of literacy for *all* adults since the Adult Performance Level Surveys conducted in the early 1970s. It is also the first in-person literacy assessment involving the prison population. A second adult literacy survey, the National Assessment of Adult Literacy (NAAL), is planned for 2003.

PERIODIC SURVEY OF A SAMPLE OF ADULTS LIVING IN HOUSEHOLDS OR PRISONS:

Assesses literacy skills:
- Prose
- Document
- Quantitative

Collects background data on:
- Demographics
- Education
- Labor Market Experiences
- Income
- Activities

Purpose

To (1) evaluate the English language literacy skills of adults (16 years and older) living in households or prisons in the United States; (2) relate the literacy skills of the nation's adults to a variety of demographic characteristics and explanatory variables; and (3) compare the results with those from the 1985 Young Adult Literacy Assessment and the 1990 Workplace Literacy Survey.

Components

The 1992 survey consisted of one component that was administered to three different representative samples: a national household sample; supplemental state household samples for 12 states (California, Florida, Illinois, Indiana, Iowa, Louisiana, New Jersey, New York, Ohio, Pennsylvania, Texas, Washington); and a national sample of federal and state prison inmates. Responses from the national, state, and prison samples were combined to yield the best possible performance estimates.

National Adult Literacy Survey. The 1992 survey assessed the literacy skills of a representative sample of the U.S. adult population using simulations of three kinds of literacy tasks that adults would ordinarily encounter in daily life (prose, document, and quantitative literacy). The data were collected through in-person interviews with adults who were living in households, or federal or state prisons. Adults were defined as individuals 16 years or older for the national and prison samples, and 16 to 64 years of age for the state samples. In addition to the cognitive tasks, the personal interview gathered information on demographic characteristics, language background, educational background, reading practices, and labor market experiences. To ensure comparability across all samples, the literacy tasks assessed were the same for all three samples. Background data varied somewhat between the household and prison samples—labor force questions were irrelevant to prisoners, and questions about criminal behavior and sentences were relevant only to prisoners.

Literacy assessment. The pool of literacy tasks used to measure adult proficiencies consisted of 165 literacy questions—41 prose, 81 document, and 43 quantitative. To ensure that valid comparisons could be made by linking the scales to those of the 1985 Young Adult Literacy Assessment, 85 tasks from that survey were included in the 1992 survey. An additional 80 new tasks were developed specifically to complement and enhance the original 85 tasks. The literacy tasks administered in NALS varied widely in terms of materials and content. The six major context/content areas were: home and family; health and safety; community and citizenship; consumer electronics; work; and leisure and recreation. Each adult was given a subset (about 45) of the total pool of assessment tasks to complete. Each of the tasks extended over a range of difficulty on the three literacy scales. The new tasks were designed to simulate the way in which people use various types of materials and to require different strategies for successful performance.

The responses to the literacy assessment were pooled and reported by proficiency scores, ranging from 0 to 500, on three separate scales, one each for prose, document, and quantitative literacy. By examining the overall characteristics of individuals who performed at each literacy level on each scale, it is possible to identify factors associated with higher or lower proficiency in reading and using prose, documents, and quantitative materials.

Background information. Background information collected for the state and household samples included data on ***background and demographics***—country of birth, languages spoken or read, access to reading materials, size of household, educational attainment of parents, age, race/ethnicity, and marital status; ***education***—highest grade completed in school, current aspirations, participation in adult education classes, and education received outside the country; ***labor market experiences***—employment status, recent labor market experiences, and occupation; ***income***—personal and household; and ***activities***—voting behavior, hours spent watching television, frequency and content of newspaper reading, and use of literacy skills for work and leisure. Respondents from each of the 12 participating states were also asked 5 state-specific questions.

To address issues of particular relevance to the prison population, a separate background questionnaire was developed for the prison sample. This instrument drew questions from the 1991 Survey of Inmates of State Correctional Facilities, sponsored by the Department of Justice's Bureau of Justice Statistics. The background questionnaire for the prison population addressed the following major topics: general and language background; educational background and experience; current offenses and criminal history; prison work assignments and labor force participation prior to incarceration; literacy activities and collaboration; and demographic information.

Periodicity

NALS was conducted in 1992. A second adult literacy study is scheduled for 2003.

2. USES OF DATA

Results from NALS provide the most detailed portrait that has ever been available on the condition of literacy in this nation and on the unrealized potential of its citizens. NALS data provide vital information to policymakers, business and labor leaders, researchers, and citizens. The survey results can be used to:

- describe the levels of literacy demonstrated by the adult population as a whole and by adults in various subgroups (e.g., those targeted at risk, prison inmates, and older adults);
- characterize adults' literacy skills in terms of demographic and background information (e.g., reading characteristics, education, and employment experiences);
- profile the literacy skills of the nation's workforce;
- compare assessment results from the current study with those from the 1985 Young Adult Literacy Assessment;

- interpret the findings in light of information-processing skills and strategies, so as to inform curriculum decisions concerning adult education and training; and
- increase understanding of the skills and knowledge associated with living in a technological society.

3. KEY CONCEPTS

Some of the key concepts related to the literacy assessment are described below. See the NALS Electronic Codebook or appendices of NALS reports for lists and descriptions of variables.

Literacy. The ability to use printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential. This definition goes beyond simply decoding and comprehending text to include a broad range of information-processing skills that adults use in accomplishing the range of tasks associated with work, home, and community contexts.

Prose Literacy. The ability to locate information contained in expository or narrative prose in the presence of related but unnecessary information, find all the information, integrate information from various parts of a passage of text, and write new information related to the text. Expository prose consists of printed information in the form of connected sentences and longer passages that define, describe, or inform, such as newspaper stories or written instructions. Narrative prose tells a story, but is less frequently used by adults in everyday life than by school children, and did not occur as often in the text presented in NALS as prose literacy tasks. Prose varies in its length, density, and structure.

Document Literacy. The ability to locate information in documents, repeat the search as many times as needed to find all the information, integrate information from various parts of a document, and write new information as requested in appropriate places in a document, while screening out related but inappropriate information. Documents differ from prose text in that they are more highly structured. Documents consist of structured prose and quantitative information in complex arrays arranged in rows and columns, such as tables, data forms, and lists (simple, nested, intersected, or combined); in hierarchical structures, such as tables of contents or indexes; or in two-dimensional visual displays of quantitative information, such as graphs, charts, and maps.

Quantitative Literacy. The ability to use quantitative information contained in prose or documents (specifically the ability to locate quantities while screening out related but unneeded information), repeat the search as many times as needed to find all the numbers, integrate information from various parts of a text or document, infer the necessary arithmetic operation(s), and perform arithmetic operation(s). Quantities can be located in either prose texts or in documents. Quantitative information may be displayed visually in graphs, maps, or charts, or it may be displayed numerically using whole numbers, fractions, decimals, percentages, or time units (hours and minutes).

Literacy Scales. Three scales used to report the results for prose, document, and quantitative literacy. These scales, each ranging from 0 to 500, are based on those established for the 1985 Young Adult Literacy Assessment. The scores on each scale represent degrees of proficiency along that particular dimension of literacy. The literacy tasks administered in the 1992 survey varied widely in terms of materials, content, and task requirements, and thus in difficulty. A careful analysis of the range of tasks along each scale provides clear evidence of an ordered set of information-processing skills and strategies along each scale. To capture this ordering, each scale was divided into five levels that reflect this progression of information-processing skills and strategies: Level 1 (0 to 225), Level 2 (226 to 275), Level 3 (276 to 325), Level 4 (326 to 375), and Level 5 (376 to 500). Level 1 comprised those adults who could consistently succeed with Level 1 literacy tasks but not with Level 2 tasks, as well as those who could not consistently succeed with Level 1 tasks and those who were not literate enough in English to take the test at all. Adults in Levels 2 through 4 were consistently able to succeed with tasks at their level but not with the next more difficult level of tasks. Adults in Level 5 were consistently able to succeed with Level 5 tasks.

Succeed Consistently. Indicates that a person at or above a given level of literacy has at least an 80 percent chance of correctly responding to a particular task. This 80 percent criterion is more stringent than the 65 percent standard used in the National Assessment of Educational Progress (NAEP—see chapter 20) for measuring what school children know and can do.

4. SURVEY DESIGN

The 1992 NALS was designed and administered by the Educational Testing Service (ETS). A subcontract was awarded to Westat, Inc. for sampling and field data collection. A committee of experts from business and industry, labor, government, research, and adult education worked with the ETS staff to develop the definition of literacy that underlies NALS, as well as to prepare the assessment objectives that guided the selection and construction of assessment tasks. In addition to this Literacy Definition Committee, a Technical Review Committee was formed to help ensure the soundness of the assessment design, the quality of the data collected, the integrity of the analyses conducted, and the appropriateness of the interpretations of the final results. The prison survey was developed in consultation with the Bureau of Justice Statistics and the Federal Bureau of Prisons. The survey design for the 1992 survey is described below.

Target Population

The target population for the national household sample consisted of adults 16 years and older in the 50 states and the District of Columbia who, at the time of the survey, resided in private households or college dormitories. The target population for the supplemental state household sample consisted of individuals 16 to 64 years of age who, at the time of the survey, resided in private households or college dormitories in the participating state (California, Florida, Illinois, Indiana, Iowa, Louisiana, New Jersey, New York, Ohio, Pennsylvania, Texas, or Washington). Individuals residing in other institutions—nursing homes, group homes, or psychiatric facilities—were not included in the household samples. The target population for the prison sample consisted of adults 16 years or older who were in state or federal prisons at the time of the survey; those held in local jails, community-based facilities, or other types of institutions were not included.

Sample Design

Because this 1992 survey was designed to provide data representative at the national level (including prison inmates) and at the state level for participating states, it included three different samples: a national household sample, supplemental state household samples for 12 states, and a supplemental national sample of state and federal prison inmates.

Household samples. The sample design for the national and state household samples involved a four-stage stratified area sample: (1) the selection of primary sampling units (PSUs) consisting of counties or contiguous groups of counties; (2) the selection of segments (within the selected PSUs) consisting of census blocks or groups of contiguous census blocks; (3) the selection of households within the segmented samples; and (4) the selection of age-eligible individuals within each selected household. The sample design requirements called for an average cluster size of seven interviews (i.e., seven completed background questionnaires per segment). In addition, a reserve sample at the household level of approximately 5 percent of the size of the main sample was selected and set aside in case of shortfalls due to unexpectedly high vacancy and nonresponse rates.

One national area sample was drawn for the national household sample, and 12 independent state-specific area samples were drawn from the 12 states participating in the supplemental state samples. The sample designs used for all 13 samples were similar, with one major difference. In the national sample, Black and Hispanic respondents were sampled at about double the rate of the remainder of the population to assure reliable estimates of their literacy proficiencies, whereas the state samples used no oversampling.

The first stage of sampling involved the selection of PSUs. A national sampling frame of 1,404 PSUs was constructed primarily from 1990 census data, stratified on the basis of region, metropolitan status, percent Black, percent Hispanic, and whenever possible, per capita income. Using this frame, 101 PSUs were selected for the national sample. The national frame of PSUs, subdivided at state boundaries if needed, was used to construct individual state frames for the supplemental state sample; a sample of 8 to 12 PSUs was selected within each of the given states. All PSUs were selected with probability proportional to the PSU's 1990 population.

The second stage of sampling involved the selection of segments within the selected PSUs. The Bureau of Census' Topologically Integrated Geographical Encoding and Referencing (TIGER) System File was used for the production of segment maps. The segments were selected with probability proportional to size where the measure of size for a segment was a function of the number of year-round housing units within the segment. The oversampling of Blacks and Hispanic respondents for the national sample was carried out at the segment level, where segments were classified as high minority (segments more than 25 percent Black or Hispanic population) or not high minority.

The third stage of sampling involved the selection of households within the segmented samples. Westat field staff visited all selected segments in the fall of 1991 and prepared lists of all housing units within the boundaries of each segment as determined by the 1990 census block maps. The lists were used to construct the sampling frame for households. Households were selected with equal probability within each segment, except for White, non-Hispanic households in high minority segments in the national sample, which were subsampled so that the sampling rates for White, non-Hispanic respondents would be about the same overall.

The fourth stage of sampling involved the selection of one or two adults within each selected household during the data collection phase of the survey. One person was selected at random from households with fewer than four eligible members; two persons were selected from households with four or more eligible members. Using a screener, the interviewer constructed a list of age-eligible household members (16 and older for the national sample; 16 to 64 for the state sample) for each selected household. The interviewers, who were instructed to list the eligible household members in descending order by age, then identified one or two household members to interview, based on computer-generated sampling messages that were attached to each questionnaire in advance.

Prison sample. There were two stages of selection for the prison sample. The first stage involved the selection of state or federal correctional facilities. The sampling frame for the correctional facilities was based on the 1990 census of federal and state prisons, updated in mid-1991. The facility frame was stratified prior to sample selection on the basis of type of facility (federal or state prison), region of country, inmate gender composition, and type of security. A sample of 88 facilities and a reserve sample of 8 facilities was then drawn from the frame based on probability proportional to size, where the measure of size for a given facility was equal to the inmate population. The second stage of sampling involved the selection of inmates within each selected facility, using a list of names obtained from the facility administrators. An average of 12 inmates were selected from each facility based on a probability inversely proportional to their facility's inmate population (up to a maximum of 22 interviews in a facility), so that the product of the first and second stage probabilities would be constant.

Assessment Design

Building on the 1985 Young Adult Literacy Assessment and the 1991 Workplace Literacy Survey, the NALS Technical Committee adopted the definition of literacy and the literacy scales—prose, document, and quantitative—used in the previous surveys. The materials were selected to represent a variety of contexts and contents: home and family; health and safety; community and citizenship; consumer electronics; work; and leisure and recreation.

BIB spiraling. The survey design gave each respondent a subset of the total pool of literacy tasks, while at the same time ensuring that each of the 165 tasks was administered to a nationally representative sample of the adult population. The design most suitable for this purpose is a variant of standard matrix sampling called balanced incomplete block (BIB) design.

Literacy tasks were assigned to blocks or sections that could be completed in about 15 minutes, and these blocks were then compiled into booklets so that each block appeared in each position (first, middle, and last) and each block was paired with every other block. Thirteen blocks of simulation tasks were assembled into 26 unique booklets, each of which contained four blocks of tasks: the core (same for all exercise booklets), and three cognitive blocks. Each booklet could be completed in about 45 minutes.

Pretests. A field test of the national household sample was conducted in the spring of 1991 using a sample of 2,000 adults drawn from 16 PSUs. The purposes of the field test were to evaluate the impact of incentives on response rates, performance, and survey costs; to evaluate newly developed literacy exercises for item bias and testing time; and to evaluate the administration and appropriateness of the background questions. As a result of the field test, some of the literacy tasks and their scoring guides were revised or dropped from the final assessment.

For the prison sample, a small pretest was conducted at the Roxbury Correctional Institution in Hagerstown, Maryland. This pretest was designed to evaluate the ease of administration of the survey instruments, survey administration time, within-facility procedures, and inmate reaction to the survey. The pretest demonstrated that several changes to the background questionnaire would facilitate administration. Administrative procedures were also refined to reflect lessons learned during the pretest.

Data Collection and Processing

The survey data were collected through in-person household or prison interviews during the first eight months of 1992. As field operations were completed, the data were shipped to ETS for processing. Further description follows.

Reference dates. Respondents answered the employment status and weekly wages questions for the week before the survey was administered.

Data collection. During January and February of 1992, field interviewers, supervisors, and editors received extensive training both in general and survey-specific interview techniques. The NALS field period began in February 1992, immediately following the completion of the first interviewer training sessions, and lasted 28 weeks, until the end of August. All three survey sample groups were worked simultaneously (except for the state of Florida where data were not collected until 1993). Except for a small, experimental "no incentive" group, all household participants who completed as much of the assessment as their skills allowed received $20 for their time. More than 400 trained interviewers visited about 44,000 households to select and interview almost 31,000 adults. In addition, over 1,147 prison inmates at 87 facilities were interviewed.

Each survey participant was asked to spend approximately one hour responding to survey questions and tasks. Data collection instruments included the screener (designed to enumerate household members and select survey respondents), the background questionnaire, and the literacy exercise booklets. Answering the screener and background questionnaire required no reading or writing skills; to ensure standardized administration, the questions on each were read to respondents in English or Spanish and the answers recorded by the assessment interviewer. Each of the exercise booklets had a corresponding interview guide, with specific instructions to the interviewer for directing the exercise booklet. Reading and writing skills in the English language were required to complete the exercise booklet. When a sampled respondent did not complete any or all of the survey instruments, the interviewer was required to complete a noninterview report form. Field supervisors reviewed the noninterview forms to determine the case's potential for conversion, and the data collected on the form were processed for nonresponse analysis.

Following the completion of an interview, interviewers edited all materials for legibility and completeness. The interviewers sent their completed work to their regional supervisors for a complete edit of the instruments, quality control procedures, and any required data retrieval. As these tasks were completed, the cases were shipped to ETS for processing.

During the data collection process, two special quality control procedures were implemented to identify any households or dwellings missed during the listing phase: the missing structure procedure and the missed dwelling unit procedure. These procedures were used to give these missed structures and dwelling units a chance of selection at time of data collection.

The field effort occurred in three overlapping stages:

(1) *Initial phase.* Each area segment was assigned by the regional supervisor to an interviewer, who followed certain rules in making a prescribed number of calls (a maximum of four was used) to every sampled dwelling in the segment.

(2) *Reassignment phase.* Cases that did not result in completed interviews during the initial phase were reviewed by the regional supervisor, and a subset was selected for reassignment to another interviewer in the same PSU or an interviewer from a nearby PSU.

(3) *Special nonresponse conversion phase.* The home office assembled a special traveling team of the most experienced or productive interviewers to perform a nonresponse conversion effort, under the supervision of a subset of the field supervisors.

Data processing. Coding and scoring staff underwent intensive training prior to the actual coding/scoring. A scoring supervisor monitored both the coding of the questionnaires and the scoring of the exercise booklets. The background questionnaire was designed to be read by a computerized scanning device. Nearly all the simulation tasks contained in the exercise booklet were open-ended; with scoring guides as examples, responses to these items were classified as correct, incorrect, or omitted by trained readers. Responses from the screener and scores from the exercise booklets were transferred to scannable answer sheets. Each survey instrument's scannable forms were batched and sent to the scanning department at regular intervals. As the different instruments were processed, the data were transferred to a database on the main ETS computer for editing.

Editing. Several quality control procedures related to data collection were used during the field operation: an interviewer field edit, a complete edit of all documents by a trained field editor, validation of 10 percent of each interviewer's closeout work, and field observation of both supervisors and interviewers. Additional edits were done during data processing. These included an assessment of the internal logic and consistency of the data received. Discrepancies were corrected whenever possible. The background questionnaires were also checked to make sure that the skip patterns had been followed and all data errors were resolved. In addition, a random set of exercise booklets was selected to provide an additional check on the accuracy of transferring information from booklets and answer sheets to the database.

Estimation Methods

Weighting was used in the 1992 NALS, prior to the calculation of base weights. Responses to the literacy tasks were scored using item response theory (IRT) scaling. A multiple imputation procedure based on plausible values methodology was used to estimate the literary proficiencies of individuals who completed literacy tasks. An innovative approach was implemented to impute missing cognitive data in order to minimize distortions in the population proficiency estimates due to nonresponse to the literacy booklet.

Weighting. Full sample and replicate weights were calculated for survey respondents who completed the exercise booklet; those who could not start the exercises because of a language barrier, a physical or mental barrier, or a reading or writing barrier; and those who refused to complete the exercises but had completed background questionnaires. Demographic variables critical to the weighting were recoded and imputed, if necessary, prior to the calculation of base weights. (See Imputation below.) Separate sets of weights were computed for the incentive and "no incentive" samples.

Household samples. A base weight was computed for each eligible record. The base weight initially was computed as the reciprocal of the product of probabilities of selection for a respondent at the PSU, segment, dwelling unit, and person levels. The final base weight included adjustments to reflect the selection of the reserve sample, the selection of missed dwelling units, and the chunking process conducted during the listing of the segments, and to account for the subsample of segments assigned to the "no incentive" experiment and the subsampling of respondents within households. The base weights for each sample were then poststratified to known 1990 census population totals, adjusted for undercount. This first-level stratification provided sampling weights with lower variation and adjusted for nonresponse. State records were poststratified separately from national records to provide a common base for applying composite weighting factors; population totals were calculated separately for each distinct group.

Composite weights were developed so that NALS data could be used to produce both state and national statistics. For the household samples, a composite weight was computed as the product of the poststratified base weight and a compositing factor which combined the national and state sample data in an optimal manner, considering the differences in sample design, sample size, and sampling error between the two sampled groups. Up to four different compositing factors were used in each of the 11 participating states, and a pseudo factor (equal to one) was used for all persons 65 and older and for all national sample records from outside the 11 participating states.

To compute the final sample weights, the composite weights were adjusted to known 1990 census counts (adjusted for undercount), using a poststratification raking ratio adjustment. The cells used for raking were defined to the finest combination of age, race/ethnicity, sex, education, and geographic indicators (e.g., MSA vs. non-MSA) that the data would allow. Raking adjustment factors were calculated separately for each of the state samples and then for the remainder of the United States.

The above steps used to create the final sample weights were repeated for 60 strategically constructed subsets of the household sample to create a set of replicate weights to be used for variance estimation using the jackknife method.

Prison sample. Base weights for the prison respondents were constructed to be equal to the reciprocal of the product of the selection probabilities for the facility and the inmate within the facility. These weights were then nonresponse-adjusted to reflect both facility and inmate nonresponse. To compute the final sample weights, the resulting nonresponse-adjusted weights were then raked to agree with independent estimates for certain subgroups of the prison population. The above procedures were repeated for 45 strategically constructed subsets of the prison sample to create a set of replicate weights to be used for variance estimation using the jackknife method.

Scaling. Since NALS used a variant of matrix sampling and since different respondents received different sets of tasks, it would be inappropriate to report its results using conventional scoring methods based on the number of correct responses. The literacy assessment results are reported using IRT scaling, which assumes some uniformity in response patterns when items require similar skills. Such uniformity can be used to characterize both examinees and items in terms of a common scale attached to the skills, even when all examinees do not take identical sets of items. Comparisons of items and examinees can then be made in reference to a scale, rather than to the percent correct. IRT scaling also allows the distributions of examinee groups to be compared.

The results of the 1992 literacy assessment are reported on three scales (prose, document, and quantitative) that were established for the 1985 Young Adult Literacy Assessment. Separate IRT linking and scaling were carried out for each of the three domains, using the three-parameter logistic (3PL) scaling model from item response theory. This is a mathematical model for estimating the probability that a particular person will respond correctly to a particular item from a single domain of items. The probability is given as a function of a parameter characterizing the proficiency of that person, and three parameters characterizing the properties of that item. Item parameters needed for the 3PL scaling model were estimated by linking each of the literacy scales used in the 1992 survey to the 1985 Young Adult Literacy Assessment scales.

Imputation. Imputation was performed prior to weighting on missing demographic items considered critical to weighting. Literacy proficiencies of respondents were estimated using a multiple imputation procedure based on plausible values methodology. Missing cognitive data were also imputed.

Demographic data. Demographic variables critical to the weighting (race/ethnicity of the head of household; sex, age, race/ethnicity, and education of the respondent) were recoded and collapsed to required levels, and imputed, if necessary, prior to the calculation of base weights. Data from the background questionnaire were preferred for all items except race/ethnicity of the head of household, which was collected on the screener. For the few cases in which the background questionnaire measure was missing, the screener measure was generally available and was used as a direct substitute. The amount of missing data remaining after substitution was small, making the imputation task fairly straightforward. A standard (random within class) hot-deck imputation procedure was performed for particular combinations of fields that were missing. Imputation flags were created for each of the five critical fields to indicate whether data were originally reported or were based on substitution or imputation. The imputed values were used only for the sample weighting process.

Literacy proficiency estimation (plausible values). A multiple imputation procedure based on plausible values methodology was used to estimate respondents' literacy proficiency in the 1992 NALS. When analyzing the distribution of proficiencies in a group of persons, more efficient estimates can be obtained from a sample design similar to that used in this 1992 survey. Such designs solicit relatively few cognitive responses from each sampled respondent but maintain a wide range of content representation when responses are summed for all respondents.

In the 1992 survey, all proficiency data were based on two types of information: responses to the background questions and responses to the cognitive items. As an intermediate step, a functional relationship between the two sets of information was calculated for the total sample, and this function was used to obtain unbiased proficiency estimates for population groups with reduced error variance. Possible values for a respondent's proficiency were sampled from a posterior distribution that is the product of two functions: the conditional distribution of proficiency given the pattern of background variables, and the likelihood function of proficiency given the pattern of responses to the cognitive items. Since exact matches of background responses are quite rare, NALS used more than 200 principal components to summarize the background information, capturing more than 99 percent of the variance. More detailed information on the plausible values methodology used in the 1992 survey is available in the *Technical Report and Data File User's Manual for the 1992 National Adult Literacy Survey* (NCES 2001–467).

Cognitive data. New procedures were implemented in the 1992 NALS to minimize distortions in the population proficiency estimates due to nonresponse to the literacy booklets. When a sampled individual decided to stop the assessment (answered less than five literacy items per scale), the interviewer used a standardized nonresponse coding procedure to record the reason why the person was stopping. This information was used to classify nonrespondents into two groups: (1) those who stopped the assessment for literacy-related reasons (e.g., language difficulty, mental disability, or reading difficulty not related to a physical disability), and (2) those who stopped for reasons unrelated to literacy (e.g., physical disability or refusal). About half of the individuals did not complete the assessment for reasons related to their literacy skills; the other respondents gave no reason for stopping, or gave reasons unrelated to their literacy.

To represent the range of implied causes of missing literacy responses, the imputation procedure selected relied on background variables and self-reported reasons for nonresponse, in addition to the functional relationship between background variables and proficiency scores for the total population. It treated "consecutively missing" data from the literacy booklet instrument differently depending on whether the nonrespondents' reasons were related or unrelated to their literacy skills: (1) those who gave literacy-related reasons were treated as wrong answers, based on the assumption that they could not have correctly completed the literacy tasks, whereas (2) those who gave no reason or cited reasons unrelated to literacy skills for not completing the assessment were essentially ignored (considered not reached), since it could not be assumed that their answers would have been either correct or incorrect. The proficiencies of such respondents were inferred from the proficiencies of other adults with similar characteristics using the plausible values methodology described above.

Future Plans

A second survey, the National Assessment of Adult Literacy (NAAL) is planned for 2003.

5. DATA QUALITY AND COMPARABILITY

The NALS sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. In addition, NCES conducted special evaluation studies to examine issues related to the quality of NALS. These studies included: (1) a study of the role of incentives in literacy survey research; (2) an evaluation of its sample design and composite estimation; and (3) an evaluation of the construct validity of the adult literacy scales.

Sampling Error

In the 1992 survey, the use of a complex sample design, adjustments for nonresponse, and poststratification procedures resulted in dependence among the observations. Therefore, a jackknife replication method was used to estimate the sampling variance. The mean square error of replicate estimates around their corresponding full sample estimate provides an estimate of the sampling variance of the statistic of interest. The replication scheme was designed to produce stable estimates of standard errors for national and prison estimates as well as for the 12 individual states.

The advantage of compositing the national and state samples during sample weighting was the increased sample size, which improved the precision of both the state and national estimates. However, biases could be present because the national PSU sample strata were not designed to maximize the efficiency of state-level estimates.

Nonsampling Error

The major source of nonsampling error in the 1992 NALS was nonresponse error; special procedures were developed to minimize potential nonresponse bias based on how much of the survey the respondent completed. Other possible sources of nonsampling error were random measurement error and systematic error due to interviewers, coders, or scorers.

Coverage error. Coverage error could result from either the sampling frame of households or prisons being incomplete or from a household's or prison's failure to include all adults 16 years and older on the lists from which the sampled respondents were drawn. Special procedures and edits were built into NALS to review both listers' and interviewers' ongoing work and to give any missed structures and/or dwelling units a chance of selection at data collection. However, just as all other household personal interview surveys have persistent undercoverage problems, the 1992 survey had problems in population coverage due to interviewers not gaining access to households in dangerous neighborhoods, locked residential apartment buildings, and gated communities.

Nonresponse error.

Unit nonresponse. Since three survey instruments—screener, background questionnaire, and exercise booklet—were required for the administration of the survey, it was possible for a household or respondent to refuse to participate at the time of the administration of any one of these instruments. Because the screener and background questionnaire were read to the survey participants in English or Spanish, but the exercise booklet required reading and writing in the English language, it was possible to complete the screener or background questionnaire but not the exercise booklet, and vice versa. Thus, response rates were calculated for each of the three instruments for the household samples. For the prison sample, there were only two points at which a respondent could not respond—at the administration of the background questionnaire or exercise booklet.

The response rate to the background questionnaire was 80.5 percent. For the household samples, the response rates exclude individuals who were not paid incentives. Also excluded are the respondents to the Florida state survey, which had a delayed administration.

The combined national and state household target sample in the 1992 NALS included 43,783 representative housing units, of which 5,405 were vacant. Approximately 89 percent of the occupied households completed a screener.

The household sample screening effort identified a total of 30,806 eligible respondents, of which 24,939 (81.0 percent unweighted) completed the background questionnaire. For the prison sample, 87 of the 88 sampled facilities participated in the survey. Of the 1,340 inmates selected, 1,147 (85.6 percent unweighted) completed the background questionnaire.

For the occupied households, "refusal or breakoff" was the most common explanation for nonresponse to the screener and background questionnaire. The second most common explanation was "not at home after maximum number of calls." Nonresponse also resulted from language, physical, and mental problems. Housing units or individuals who refused to participate before any information was collected about them, or who did not answer a sufficient number of background questions, were never incorporated into the database. Because these individuals were unlikely to know that the survey intended to assess their literacy, it was assumed that their reason for not completing the survey was not related to their level of literacy.

Literacy assessment booklets were considered complete if at least five items were answered on each scale. A total of 24,944 household sample members were classified as eligible for the exercise booklet. Of these, 88.6 percent completed the booklet and another 6.1 percent partially completed the exercise. Of the 1,147 eligibles in the prison sample, 86.8 percent completed the booklet and another 9.3 percent partially completed it.

There were reasons to believe that the literacy performance data were missing more often for adults with lower levels of literacy than for adults with higher levels. Field test evidence and experience with surveys indicated that adults with lower levels of literacy were more likely than adults with higher proficiencies either to decline to respond to the survey at all or to begin the assessment but not complete it. Ignoring this pattern of missing data would have resulted in overestimating the literacy skills of adults in the United States. Therefore, to minimize bias in the proficiency estimates due to nonresponse to the literacy assessment, special procedures were developed to impute the literacy proficiencies of nonrespondents who completed fewer than five literacy tasks.

Item nonresponse. For each background questionnaire, staff verified that certain questions providing critical information for weighting and data analyses had been answered, namely education level, employment status, parents' level of education, race, and sex. If a response was missing, the case was returned to the field for data retrieval. Therefore, item response rates for completed background questionnaires were quite high, although they varied by type of question. Questions asking country of origin (first question in the booklet) and sex (last question in the booklet) had nearly 100 percent response rates, indicating that most respondents attempted to complete the entire questionnaire. Response rates were lower, however, for questions about income and educational background.

The electronic codebook provides counts of item nonresponse. These, however, have to be considered in terms of the number of adults that were offered each task, because a great deal of the missing data is missing by design.

Measurement error. All background questions and literacy tasks underwent extensive review by subject area and measurement specialists, as well as scrutiny to eliminate any bias or lack of sensitivity to particular groups. Special care was taken to include materials and tasks that were relevant to adults of widely varying ages. During the test development stage, the tasks were submitted to test specialists for review, part of that involved checking the accuracy and completeness of the scoring guide. After preliminary versions of the assessment instruments were developed and after the field test was conducted, the literacy tasks were closely analyzed for bias or "differential item functioning." The goal was to identify any assessment tasks that were likely to underestimate the proficiencies of a particular subpopulation, whether it be older adults, females, or Black or Hispanic adults. Any assessment item that appeared to be biased against a subgroup was excluded from the final survey. The coding and scoring guides also underwent further revisions after the first responses were received from the main data collection.

Interviewer error checks. Several quality control procedures related to data collection were used during the field operation: an interviewer field edit, a complete edit of all documents by a trained field editor, validation of 10 percent of each interviewer's closeout work, and field observation of both supervisors and interviewers.

Coding/scoring error checks. In order to monitor the accuracy of coding, the questions dealing with country of birth, language, wages, and date of birth were checked in 10 percent of the questionnaires by a second coder. For the industry and occupation questions, 100 percent of the questionnaires were recoded by a second coder. Twenty percent of all the exercise booklets were subjected to a reader reliability check, which entailed a scoring by a second reader. There was a high degree of reader reliability across tasks—ranging from 88.1 to 99.9 percent—with an average agreement of 97 percent. For 133 out of 165 open-ended tasks, the agreement between the two readers was above 95 percent.

Data Comparability

One of the major goals of this survey was to compare its results to the 1985 Young Adult Literacy Assessment and other large assessment studies.

Comparisons with the 1985 Young Adult Literacy Assessment. Comparisons are possible because the sample design, item pool, and methodology used in the 1985 Young Adult Literacy Assessment and the 1992 survey were very similar. Literacy tasks for each survey were developed using the same definition of literacy, and a subset of identical tasks was administered in both assessments. Scoring guides were the same for both surveys. Both gave nearly identical incentive payments to participants ($15 in 1985 and $20 in 1992). The literacy scales used in the two surveys were linked so that the scores could be reported on a common scale.

Nevertheless, there were some differences in procedures for the two surveys. For example, missing responses to the literacy tasks were handled differently. In the 1985 Young Adult Literacy Assessment, individuals who could not answer six core literacy tasks and those who spoke only Spanish were excluded from the analyses. In the 1992 survey, however, a special procedure was used to impute literacy proficiencies for literacy-related nonrespondents.

Due to such procedural differences, direct comparisons of the results of the two surveys are not simple and straightforward. However, because the 1992 sample is more inclusive than the 1985 sample, subsamples that have more exact counterparts in the 1985 survey can be selected. For instance, the initial report from the 1992 NALS presented data, using no subsample matching, indicated that young adults in 1992 were somewhat less literate than their predecessors in 1985. However, when a comparison was made between matched subsamples of the 1985 and 1992 survey respondents based on reasons for nonresponse, the proficiency differences decreased significantly. Furthermore, results from partition analysis of the two surveys' matched subsamples—based on change due to variations in demographic characteristics versus change not related to demography—suggest that most of the observed declines in the average literacy skills of young adults over time can be accounted for by shifts in the composition of the population and by changes across the assessments in the rules used to include or exclude nonrespondents.

Comparisons with the 1993 GED. Comparisons between NALS and GED examinees are explored in *The Literacy Proficiencies of GED Examinees: Results from the GED-NALS Comparison Study* (by Janet Baldwin, Irwin S. Kirsch, Don Rock, and Kentaro Yamamoto; American Council on Education and Educational Testing Service: 1993). The GED Tests and NALS instruments have a considerable degree of overlap in what they measure. Both assess skills that appear to represent verbal comprehension and reasoning, or the ability to understand, analyze, interpret, and evaluate written information and apply fundamental principles and concepts. Despite the considerable degree of overlap, the two instruments also measure somewhat different skills. For example, the GED Tests seem to tap unique dimensions of writing mechanics and mathematics, while the adult literacy scales appear to tap unique dimensions of document literacy. In addition, the evidence shows that there are no differences in the average prose, document, or quantitative literacy skills of those adults who terminated their schooling at the high school or GED level.

6. CONTACT INFORMATION

For content information on the National Adult Assessments of Literacy, contact:

Andrew J. Kolstad
Phone: (202) 502–7374
E-mail: andrew.kolstad@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

Adult Literacy in America: A First Look at the Findings of the National Adult Literacy Survey, NCES 93–275, by I.S. Kirsch, A. Jungeblut, and L. Jenkins. Washington, DC: 1993.

Technical Report and Data File User's Manual for the 1992 National Adult Literacy Survey, NCES 2001–457, by I. Kirsch, K. Yamamoto, N. Norris, D. Rock, A. Jungeblut, P. O'Reilly, A. Campbell, L. Jenkins, A. Kolstad, M. Berlin, L. Mohadjer, J. Waksberg, H. Goksel, J. Burke, S. Rieger, J. Green, M. Klein, P. Mosenthal, and S. Baldi. Washington, DC: 2000.

Survey Design

Assessing Literacy: The Framework for the National Adult Literacy Survey, NCES 92–113, by A. Campbell and I.S. Kirsch. Washington, DC: 1992.

Chapter 24: International Adult Literacy Survey (IALS)

1. OVERVIEW

The 1994 International Adult Literacy Survey (IALS) represented a first attempt to assess the literacy skills of entire adult populations in a framework that provided data comparable across cultures and languages. This collaborative project was designed to inform both education and labor market policy and program development activities in participating countries. The international portion of the study was carried out under the auspices of an International Steering Committee chaired by Canada, with each participating country holding a seat on the committee along with representatives from the Organization for Economic Cooperation and Development (OECD), European communities, and the United Nations Educational, Scientific and Cultural Organization.

1994 INTERNATIONAL STUDY OF ADULT LITERACY

IALS collected:

- Background Assessments
- Literacy Assessments

In the United States, IALS is the fourth assessment of adult literacy funded by the federal government and conducted by the Educational Testing Service (ETS). The three previous efforts were: (1) the 1992 National Adult Literacy Survey (see chapter 23); (2) the Department of Labor's (DOL) 1990 Workplace Literacy Survey; and (3) the 1985 Young Adult Literacy Survey (funded as an adjunct to the National Assessment of Educational Progress—see chapter 20). In order to maximize the comparability of estimates across countries, the IALS study chose to adopt the National Adult Literacy Survey methodology and scales. Literacy was defined along three dimensions—prose, document, and quantitative. These were designed to capture an ordered set of information-processing skills and strategies that adults use to accomplish a diverse range of literacy tasks encountered in everyday life. The background data collected in IALS provide a context for understanding the ways in which various characteristics are associated with demonstrated literacy skills.

IALS was originally conducted in seven countries (Canada, Germany, the Netherlands, Poland, Sweden, French- and German-speaking Switzerland, and the United States). A second phase was subsequently conducted in five additional countries (Australia, Flemish-speaking Belgium, Great Britain, New Zealand, and Ireland), and in a final phase included an additional 10 countries. This chapter will focus on the first phase, in which the United States participated.

Purpose

To (1) develop scales that would permit comparisons of the literacy performance of adults (16 and older) with a wide range of abilities; (2) if such an assessment could be created, describe and compare the demonstrated literacy skills of adults in different countries.

Components

Each IALS country was given a set of model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments. IALS instruments consisted of three parts: (1) a background questionnaire, which collected demographic information about respondents; (2) a set of core literacy tasks, which screened out respondents with very limited literacy skills; and (3) a main booklet of literacy tasks, used to calibrate literacy levels.

Background Questionnaire. The background questionnaire collected information on languages spoken or read; parents' educational attainment and employment; labor force experiences—employment status, recent labor force experiences, and occupation; reading and writing at work and looking for work; participation in adult education classes—courses taken, financial support, purpose; reading and writing in daily life (excluding work or school); family literacy—children's reading habits, the household's access to reading materials, hours spent watching television; and household information—total income and sources of income. The background questionnaire was to be administered in about 20 minutes.

Literacy Assessment—Core Literacy Tasks and Main Literacy Tasks. One hundred and fourteen tasks were grouped into three scales and divided into seven blocks (labeled A through G), which in turn were compiled into seven test booklets (numbered 1 through 7). Each booklet contained three blocks of tasks and was designed to take about 45 minutes to complete. Respondents began the cognitive part of the assessment by performing a set of six "core" tasks. Only those who were able to perform at least two of the six core tasks correctly (93 percent of respondents) were given the full assessment.

Periodicity

The first phase of data collection for the IALS was conducted during the autumn of 1994 in Canada, Germany, the Netherlands, Poland, Sweden, Switzerland (French and German-speaking cantons), and the United States. Data were collected from a second group of countries—Australia, Flemish-speaking Belgium, Great Britain, New Zealand, and Ireland—in 1995–96. Data were collected from a third group of countries in 1997–98. No second administration is planned.

2. USES OF DATA

IALS is designed to inform both educational and labor market policy and program development activities in participating countries. The primary objectives of the study are:

- To shed light on the relationship between microeconomic variables—such as individual literacy, educational attainment, labor market participation and employment, and macroeconomic issues—such as competitiveness, growth, and restructuring;
- To identify subpopulations that are economically and socially disadvantaged by their literacy skill profiles; and
- To establish the comparability of assessments of adult literacy.

IALS data provide comparable information about the activities and outcomes of educational systems and institutions in participating countries. Such data can lead to improvements in accountability and policymaking. These data are increasingly relevant to policy formation due to the growing political, economic, and cultural ties between countries.

3. KEY CONCEPTS

Some of the key concepts related to the IALS literacy assessment are described below.

Literacy. The ability to use printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and potential.

Prose Literacy. The ability to read and use texts of varying levels of difficulty which are presented in sentence and paragraph form, including editorials, news stories, poems, and fiction.

Document Literacy. The knowledge and skills required to locate and use information contained in formats such as job applications, payroll forms, transportation schedules, maps, tables, and graphics.

Quantitative Literacy. The knowledge and skills required to apply arithmetic operations, either alone or sequentially, to numbers embedded in printed materials, such as balancing a checkbook, calculating a tip, completing an order form, or determining the amount of interest on a loan from an advertisement.

Literacy Scales. The three scales used to report the results for prose, document, and quantitative literacy. These scales, each ranging from 0 to 500, are based on those established for the Young Adult Literacy Survey, the DOL's Workplace Literacy Survey, and the National Adult Literacy Survey. The scores on each scale represent degrees of proficiency along that particular dimension of literacy. The scales make it possible not only to summarize the literacy proficiencies of the total population and of various subpopulations, but also to determine the relative difficulty of the literacy tasks administered in IALS.

The literacy tasks administered in IALS varied widely in terms of materials, content, and task requirements, and thus in difficulty. A careful analysis of the range of tasks along each scale provides clear evidence of an ordered set of information-processing skills and strategies along each scale. To capture this ordering, each scale was divided into five levels that reflect this progression of information-processing skills and strategies: Level 1 (0 to 225), Level 2 (226 to 275), Level 3 (276 to 325), Level 4 (326 to 375), and Level 5 (376 to 500). Level 1 comprised those adults who could consistently succeed with Level 1 literacy tasks but not with Level 2 tasks, as well as those who could not consistently succeed with Level 1 tasks and those who were not literate enough to take the test at all. Adults in Levels 2 through 4 were consistently able to succeed with tasks at their level but not with the next more difficult level of tasks. Adults in Level 5 were consistently able to succeed with Level 5 tasks. The use of three parallel literacy scales makes it possible to profile and compare the various types and levels of literacy demonstrated by adults in different countries and by subgroups within those countries.

4. SURVEY DESIGN

Statistics Canada and ETS, a private testing organization in the United States, coordinated the development and management of IALS. These organizations were assisted by national research teams from the participating countries in developing the survey design. The survey design for the 1994 IALS is described below.

Target Population

The IALS target population was the civilian, noninstitutionalized population aged 16 to 65 in each country; however, countries were also permitted to sample older adults, and several did so. All IALS samples excluded full-time members of the military and people residing in institutions such as prisons, hospitals, and psychiatric facilities.

For the United States, the target population consisted specifically of civilian noninstitutionalized residents aged 16 to 65 years in the 50 states and the District of Columbia, excluding members of the armed forces on active duty, those residing outside the United States, and those with no fixed household address (i.e., the homeless or residents of institutional group quarters such as prisons and hospitals).

Sample Design

IALS was designed to provide data representative at the national level. Each country that participated in IALS agreed to draw a probability sample that would accurately represent its civilian, noninstitutionalized population aged 16 to 65. The final IALS sample design criteria specified that each country's sample should result in at least 1,000 respondents, the minimum sample size needed to produce reliable literacy proficiency estimates. Given the different sizes of the population of persons aged 16 to 65 in the countries involved, sample sizes varied considerably from country to country (ranging from 1,500 to 8,000 per country), but sample sizes were sufficiently large in all cases to support the estimation of reliable IRT item parameters.

IALS countries were strongly encouraged to select high-quality probability samples because the use of probability designs would make it possible to produce unbiased estimates for individual countries and to compare these estimates across the countries. Because the available data sources and resources were different in each of the participating countries, however, no single sampling methodology was imposed. Each IALS country created its own sample design. All countries used probability sampling for at least some stages of their sample designs, and some used probability sampling for all stages of sampling. Sampling designs were approved by expert review.

The sample for the United States was selected from a sample of individuals in housing units who were completing their final round of interviews for the Current Population Survey (CPS) in March, April, May, and June 1994. These housing units were included in the CPS for their initial interviews in December 1992 and January, February, and March 1993. The CPS is a large-scale continuous household survey of the civilian noninstitutionalized population aged 15 and over.

The sample was selected from housing units undergoing their final CPS interviews in March–June, 1994. The frame for the CPS consisted of 1990 Decennial Census files, which are continually updated for new residential construction and are adjusted for undercount, births, deaths, immigration, emigration, and changes in the armed forces.

The CPS sample is selected using a stratified multistage design. Housing units that existed at the time of the 1990 Population Census were sampled from the Census list of addresses. Housing units that did not exist at that time were sampled from lists of new construction when available and otherwise by area sampling methods. Occupants of housing units that came into existence between the time of the CPS sample selection and the time of the IALS fieldwork had no chance of being selected for IALS.

The IALS sample was confined to 60 of the 729 CPS primary sampling units (PSUs). Within these 60 PSUs, all persons aged 16 to 65 years of age in the sampled housing units were classified into 20 cells defined by race/ethnicity and education. Within each cell, persons were selected for IALS with probability proportional to their CPS weights, with the aim of producing an equal probability sample of persons within cells. A total of 4,901 persons was selected for IALS. IALS interviews were conducted in October and November 1994.

Assessment Design

The success of IALS depended on the development and standardized application of a common set of survey instruments. The test framework explicitly followed the precedent set by the National Adult Literacy Survey, basing the test on United States definitions of literacy along three dimensions—prose literacy, document literacy, and quantitative literacy—but extending the instruments into an international context. Study managers from each participating country were encouraged to submit materials such as news articles and documents that could be used to create tasks with the goal of building a new pool of literacy tasks that could be linked to established scales. IALS team field tested 175 tasks and identified 114 that were valid across cultures. Approximately half of these tasks were based on materials from outside North America. (However, each respondent was administered only a fraction of the pool of tasks, using a variant of matrix sampling.)

Each IALS country was given a set of model administration manuals and survey instruments as well as graphic files containing the pool of IALS literacy items with instructions to modify each item by translating the English text to its own language without altering the graphic representation. Certain rules governed the item modification process. For instance, some items required respondents to perform a task that was facilitated by the use of keywords. The keyword in the question might be identical, similar but not exactly the same, or a synonym of the word used in the body of the item, or respondents might be asked to choose among multiple keywords in the body of the item, only one of which was correct. Countries were required to preserve these conceptual associations during the translation process. Particular conventions used in the items—for example, currency units, date formats, and decimal delimiters—were adapted as appropriate for each country.

To ensure that the adaptation process did not compromise the psychometric integrity of the items, each country's test booklets were carefully reviewed for errors of adaptation. Countries were required to correct all errors found. However, this review was imperfect in two important respects. First, it is clear that countries chose not to incorporate a number of changes that were identified during the course of the review, believing that they "knew better." Second, the availability of empirical data from the study has permitted the identification of several additional sources of task and item difficulty that were not included in the original framework, which was based on research by Irwin Kirsch of ETS and Peter Mosenthal of Syracuse University. (See "Exploring Document Literacy: Variables Underlying the Performance of Young Adults," by I.S. Kirsch and P.B. Mosenthal, in *Reading Research Quarterly 25*: 5–30.) Item adaptation guidelines and item review procedures associated with subsequent rounds of IALS data collection were adapted to reflect this additional information.

The model background questionnaires contained two sets of questions: mandatory questions, which all countries were required to include; and optional questions, which were recommended but not required. Countries were not required to field literal translations of the mandatory questions, but were asked to respect the conceptual intent of each question in adapting it for use. Countries were permitted to add questions to their background questionnaires if the additional burden on respondents would not reduce response rates. Statistics Canada reviewed all background questionnaires except Sweden's before the pilot survey and offered comments and suggestions to each country.

Data Collection and Processing

IALS data for the first round of countries were collected through in-person household interviews in the fall of 1994. Each country mapped its national dataset into a highly structured, standardized record layout which it sent to Statistics Canada. Further description follows.

Reference dates. Respondents answered questions about jobs they may have held in the 12 months before the survey was administered.

Data collection. Statistics Canada and ETS coordinated the development and management of IALS. Participating countries were given model administration manuals and survey instruments as well as guidelines for adapting and translating the survey instruments and for handling nonresponse codings.

Countries were permitted to adapt these models to their own national data collection systems, but they were required to retain a number of key features: (1) respondents were to complete the core and main test booklets alone, in their homes, without help from another person or from a calculator; (2) respondents were not to be given monetary incentives for participating; (3) despite the prohibition on monetary incentives, interviewers were provided with procedures to maximize the number of completed background questionnaires, and were to use a common set of coding specifications to deal with nonresponse. This last requirement was critical. Because noncompletion of the core and main task booklets was correlated with ability, background information about nonrespondents was needed in order to impute cognitive data for these persons.

IALS countries were instructed to obtain at least a background questionnaire from sampled individuals. All countries participating in IALS instructed interviewers to make callbacks at households that were difficult to contact.

In general, the survey was carried out in the national language. In Canada, respondents were given a choice of English or French, and in Switzerland, samples drawn from French-speaking and German-speaking cantons were required to respond in those respective languages. When respondents could not speak the designated language, attempts were made to complete the background questionnaire so that their literacy level could be estimated and the possibility of distorted results would be reduced. In the United States, the test was given in English, but a Spanish version of the background questionnaire and bilingual interviewers were available to assist individuals whose native language was not English.

Survey respondents spent approximately 20 minutes answering a common set of background questions concerning their demographic characteristics, educational experiences, labor market experiences, and literacy-related activities. Responses to these background questions made it possible to summarize the survey results using an array of descriptive variables, and also increased the accuracy of the proficiency estimates for various subpopulations. After answering the background questions, the remainder of respondents' time was spent completing a booklet of literacy tasks designed to measure their prose, document, and quantitative skills. Most of these tasks were open-ended, requiring respondents to provide a written answer.

In the United States, the IALS interview period was from October to November 1994. IALS was conducted by 149 Census Bureau interviewers. All of them had at least 5 days of interviewer training. They were given a 1-day training on IALS and were provided with substantial training and reference materials based on the Canadian training package. They also performed a day of field training under the supervision of a regional office supervisor. Each interviewer had an average workload of 33 interviews, and the average number of response interviews per interviewer was 21. They were supervised by six regional supervisors who reviewed and commented on their work.

Before data collection, a letter was sent to the selected addresses describing the upcoming survey. The survey was limited to 90 minutes. If a respondent took more than 20 minutes per block, the interviewer was instructed to move the respondent on to the next block.

Data processing. As a condition of their participation in IALS, countries were required to capture and process their files using procedures that ensured logical consistency and acceptable levels of data capture error. Specifically, countries were advised to conduct complete verification of the captured scores (i.e., enter each record twice) in order to minimize error rates. One hundred percent keystroke validation was needed. Specific details about scoring are provided in a separate section below.

To create a workable comparative analysis, each IALS country was required to map its national dataset into a highly structured, standardized record layout. In addition to specifying the position, format, and length of each field, this International Record Layout included a description of each variable and indicated the categories and codes to be provided for that variable. Upon receiving a country's file, Statistics Canada performed a series of range checks to ensure compliance to the prescribed format. When anomalies were detected, countries corrected the problems and submitted new files. Statistics Canada did not, however, perform any logic or flow edits, as it was assumed that participating countries performed this step themselves.

Editing. Most countries followed IALS guidelines, verifying 100 percent of their data capture operation. The two countries that did not comply with this recommendation conducted sample verifications, one country at 20 percent and the other at 10 percent. Each country coded and edited its own data, mapping its national dataset into the detailed International Record Layout, which included a description of each variable and indicated the categories and codes to be provided for that variable. Industry, occupation, and education were coded using the standard international coding schemes: the International Standard Industrial Classification (ISIC), the International Standard Occupational Classification (ISOC), and the International Standard Classification of Education (ISCED). Coding schemes were provided for open-ended items; the coding schemes came with specific instructions so that coding error could be contained to acceptable levels.

Scoring. Respondents' literacy proficiencies were estimated based on their performance on the cognitive tasks administered in the assessment. Because the open-ended items used in IALS elicited a large variety of responses, responses had to be grouped in order to summarize the performance results. As they were scored, responses to IALS open-ended items were classified as correct, incorrect, or omitted. The models employed to estimate ability and difficulty were predicated on the assumption that the scoring rubrics developed for the assessment were applied in a consistent fashion within and between countries. To reinforce the importance of consistent scoring, a meeting of national study managers and chief scorers was held prior to the commencement of scoring for the main study. The group spent 2 days reviewing the scoring rubrics for all the survey items. Where this review uncovered ambiguities and situations not covered by the guides, clarifications were agreed to collectively, and these clarifications were then incorporated into the final rubrics. To provide ongoing support during the scoring process, Statistics Canada and ETS maintained a joint scoring hotline. Any scoring problems encountered by chief scorers were resolved by this group, and decisions were forwarded to all national study managers. Study managers conducted intensive scoring training using the scoring manual and discussed unusual responses with scorers. They also offered additional training to some scorers, as needed, to raise their accuracy to the level achieved by other scorers.

To maintain coding quality within acceptable levels of error, each country undertook to rescore a minimum of 10 percent of all assessments. Where significant problems were encountered, larger samples of a particular scorer's work were to be reviewed and, where necessary, their entire assignments rescored. Countries were not required to resolve contradictory scores in the main survey (as they had been in the pilot), since outgoing agreement rates were far above minimum acceptable tolerances.

Since there could still be significant differences in the consistency of scoring between countries, countries agreed to exchange at least 300 randomly selected booklets with another country sharing the same test language. In all cases where serious discrepancies were identified, countries were required to rescore entire items or discrepant code pairs.

Intra-country rescoring. A variable sampling ratio procedure was set up to monitor scoring accuracy. At the beginning of scoring, almost all responses were rescored to identify inaccurate scorers and to detect unique or difficult responses that were not covered in the scoring manual. After a satisfactory level of accuracy was achieved, the rescoring ratio was dropped to a maintenance level to monitor the accuracy of all scorers. Average agreements were calculated across all items. Precautions were taken to ensure that the first and second scores were truly independent.

Intercountry rescoring. To determine intercountry scoring reliabilities for each item, the responses of a subset of examinees were scored by two separate groups. Usually, these scoring groups were from different countries. Intercountry score reliabilities were calculated by Statistics Canada, then evaluated by ETS. Based on the evaluation, every country was required to introduce a few minor changes in scoring procedures. In some cases, ambiguous instructions in the scoring manual were found to be causing erroneous interpretations and therefore lower reliabilities.

Using the intercountry score reliabilities, researchers could identify poorly constructed items, ambiguous scoring criteria, erroneous translations of items or scoring crite-

BEST COPY AVAILABLE

ria, erroneous printing of items or scoring criteria, scorer inaccuracies, and, most important, situations in which one country consistently scored differently from another. In the latter circumstance, scorers in one country may consistently rate a certain response as being correct while those in another country score the same response as incorrect. ETS and Statistics Canada examined scoring carefully to identify situations in which scorers in one country were consistently rating a certain response as being correct while those in another country were scoring the same response as incorrect. Where a systematic error was identified in a particular country, the original scores for that item were corrected for the entire sample.

Estimation Methods

Weighting was used in the 1994 IALS to adjust for sampling and nonresponse. Responses to the literacy tasks were scored using IRT scaling. A multiple imputation procedure based on plausible values methodology was used to estimate the literacy proficiencies of individuals who completed literacy tasks.

Weighting. IALS countries used different methods for weighting their samples. Countries with known probabilities of selection could calculate a base weight using the probability of selection. To adjust for unit nonresponse, all countries poststratified their data to known population counts, and a comparison of the distribution of the age and sex characteristics of the actual and weighted samples indicates that the samples were comparable to the overall populations of IALS countries. Another commonly used approach was to weight survey data to adjust the rough estimates produced by the sample to match known population counts from sources external to IALS. This "benchmarking" procedure assumes that the characteristics of nonrespondents are similar to those of respondents. It is most effective when the variables used for benchmarking are strongly correlated with the characteristic of interest—in this case, literacy levels. For IALS, the key benchmarking variables were age, employment status, and education. All of IALS countries benchmarked to at least one of these variables. The United States used education.

Weights for the United States IALS included two components. The first assigned weights to CPS respondents, and the second assigned weights to IALS respondents.

The CPS weighting scheme was a complex one involving three components: basic weighting, noninterview adjustment, and ratio adjustment. The basic weighting compensated for unequal selection probabilities. The noninterview adjustment compensated for nonresponse within weighting cells created by clusters of PSUs of similar size; Metropolitan Statistical Area (MSA) clusters are subdivided into central city areas, and the balance of the MSA and non-MSA clusters are divided into urban and rural areas. The ratio adjustment made the weighted sample distributions conform to known distributions on such characteristics as age, race, Spanish origin, sex, and residence.

The weights of persons sampled for IALS were adjusted to compensate for the use of the four rotation groups, the sampling of the 60 PSUs, and the sampling of persons within the 60 PSUs. The IALS noninterview adjustment compensated for sampled persons for whom no information was obtained because they were absent, refused to participate, had a short-term illness, had moved or had experienced an unusual circumstance that prevented them from being interviewed. Finally, the IALS ratio adjustment ensured that the weighted sample distributions across a number of education groups conformed to March 1994 CPS estimates of these numbers.

Scaling (item response theory). The scaling model used in IALS was the two-parameter logistic model from item response theory.

Items developed for IALS were based on the framework used in three previous large-scale assessments: the Young Adult Literacy Survey (YALS), the DOL survey, and the National Adult Literacy Survey. As a result, IALS items shared the same characteristics as the items in these earlier surveys. The English version of IALS items were reviewed and tested to determine whether they fit into the literacy scales in accordance with the theory and whether they were consistent with the National Adult Literacy Survey data. Quality control procedures for item translation, scoring, and scaling followed the same procedures used in the National Adult Literacy Survey and extended the methods used in other international studies.

Identical item calibration procedures were carried out separately for each of the three literacy scales: prose, document, and quantitative literacy. Using a modified version of Mislevy and Bock's 1982 BILOG computer program—see *BILOG: Item analysis and test scoring with binary logistic models*, Scientific Software—the two-parameter logistic IRT model was fit to each item using sample weights. BILOG procedures are based on an extension of the marginal-maximum-likelihood approach described by Bock and Aitkin in their 1981 *Psychometrika* article, "Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm."

Most of the items administered in IALS were successful from a psychometric standpoint. However, despite stringent efforts at quality control, some of the assessment items did not meet the criteria for inclusion in the final tabulation of results. Specifically, in carrying out the IRT modeling used to create the three literacy scales, researchers found that a number of assessment items had significantly different item parameters across IALS countries.

Imputation. A respondent had to complete the background questionnaire, pass the core block of literacy tasks, and attempt at least five tasks per literacy scale in order for researchers to be able to estimate his or her literacy skills directly. Literacy proficiency data were imputed for individuals who failed or refused to perform the core literacy tasks and for those who passed the core block but did not attempt at least five tasks per literacy scale. Because the model used to impute literacy estimates for nonrespondents relied on a full set of responses to the background questions, IALS countries were instructed to obtain at least a background questionnaire from sampled individuals. IALS countries were also given a detailed nonresponse classification to use in the survey.

Literacy proficiencies of respondents were estimated using a multiple imputation procedure based on plausible values methodology. Special procedures were used to impute missing cognitive data.

Literary proficiency estimation (plausible values). A multiple imputation procedure based on plausible values methodology was used to estimate respondents' literacy proficiency in the 1994 IALS. When a sampled individual decided to stop the assessment, the interviewer used a standardized nonresponse coding procedure to record the reason why the person was stopping. This information was used to classify nonrespondents into two groups: (1) those who stopped the assessment for literacy-related reasons (e.g., language difficulty, mental disability, or reading difficulty not related to a physical disability); and (2) those who stopped for reasons unrelated to literacy (e.g., physical disability or refusal). About 45 percent of the individuals did not complete the assessment for reasons related to their literacy skills; the other respondents gave no reason for stopping, or gave reasons unrelated to their literacy.

When individuals cited a literacy-related reason for not completing the cognitive items, this implies that they were unable to respond to the items. On the other hand, citing reasons unrelated to literacy implies nothing about a person's literacy proficiency. Based on these interpretations, IALS adapted a procedure originally developed for the National Adult Literacy Survey to treat cases in which an individual responded to fewer than five items per literacy scale, as follows: (1) if the individual cited a literacy-related reason for not completing the assessment, then all consecutively missing responses at the end of the block of items were treated as wrong; and (2) if the individual cited reasons unrelated to literacy for not completing the assessment, then all consecutively missing responses at the end of a block were treated as "not reached."

Proficiency values were estimated based on respondents' answers to the background questions and the cognitive items. As an intermediate step, the functional relationship between these two sets of information was calculated, and this function was used to obtain unbiased proficiency estimates with reduced error variance. A respondent's proficiency was calculated from a posterior distribution that was the multiple of two functions: a conditional distribution of proficiency, given responses to the background questions; and a likelihood function of proficiency, given responses to the cognitive items.

Recent Changes

Since IALS was a onetime assessment, there are no changes to report.

Future Plans

There are no plans to conduct IALS again. However, a new survey called the International Study of Adults (ISA, also known as ALL) is being administered in 2003. The aspects of this survey that address literacy build on methodologies used in IALS.

5. DATA QUALITY AND COMPARABILITY

The literacy tasks contained in IALS and the adults asked to participate in the survey were samples drawn from their respective universes. As such, they were subject to some measurable degree of uncertainty. IALS implemented procedures to minimize both sampling and nonsampling errors. The IALS sampling design and weighting procedures assured that participants' responses could be generalized to the population of interest. Scientific procedures employed in the study design and the scaling of literacy tasks permitted a high degree of confidence in the resulting estimates of task difficulty. Quality control activities continued during interviewer training, data collection, and processing of the survey data.

In addition, special evaluation studies were conducted to examine issues related to the quality of IALS. These studies included: (1) an external evaluation of IALS methodology; (2) an examination of how similar or different the sampled persons were from the overall population; (3) an evaluation of the extent to which the literacy levels of the population in the database for each nation were predictable based on demographic characteristics; (4) an examination of the assumption of unidimensionality; and (5) an evaluation of the construct validity of the adult literacy scales.

Sampling Error

Because IALS employed probability sampling, the results were subject to sampling error. Although small, this error was rather higher in IALS than in most studies because the cost of surveying adults in their homes is so high. Most countries simply could not afford large sample sizes.

Each country provided a set of replicate weights for use in a Jackknife variance estimation procedure.

There were three situations in which nonprobability-based sampling methods were used: France and Germany used "random route" procedures for selecting households into their samples, and Switzerland used an alphabetic sort to select one member of each household. However, based on the available evidence, it is not believed that these practices introduced significant bias into the survey estimates.

In 1998, the UK Office of National Statistics coordinated the European Adult Literacy Review, a split-sample survey intended, in part, to measure the effects of sampling methods on the IALS results. This follow-up survey compared an IALS sample design with an alternative, standardized "best practice" design. Although certain differences were noted between the two samples, the IALS sample design was not confirmed to be inferior to the "best practice" design.

Nonsampling Error

The key sources of nonsampling error in the 1994 IALS were differential coverage across countries and nonresponse bias, which occurred when different groups of sampled individuals failed to participate in the survey. Other potential sources of nonsampling error included deviations from prescribed data collection procedures, and errors of logic which resulted from mapping idiosyncratic national data into a rigid international format. Scoring error, associated with scoring open-ended tasks reliably within and between countries, also occurred. Finally, because IALS data were collected and processed independently by the various countries, the study was subject to uneven levels of commonplace data capture, data processing, and coding errors.

Three studies were conducted to examine the possibility of nonresponse bias. Because the sampling frames for Canada and the United States contained information about the characteristics of sampled individuals, it was possible to compare the characteristics of respondents and nonrespondents, particularly with respect to literacy skill profiles. The Swedish National Study Team also commissioned a nonresponse follow-up study.

Coverage error. The design specifications for IALS stated that in each country the study should cover the civilian, noninstitutional population aged 16–65. It is the usual practice to exclude the institutional population from national surveys because of the difficulties in conducting interviews in institutional settings. Similarly, it is not uncommon to exclude certain other parts of a country's population that pose difficult survey problems (e.g., persons living in sparsely populated areas). The intended coverage of the surveys generally conformed well to the design specifications: each of IALS countries attained a high level of population coverage, ranging from a low of 89 percent in Switzerland to 99 percent in the Netherlands and Poland. However, it should be noted that actual coverage is generally lower than the intended coverage because of deficiencies in sampling frames and sampling frame construction (e.g., failures to list some households and some adults within listed households). In the United States, for example, comparing population sizes estimated from the survey with external benchmark figures suggests that the overall coverage rate for the CPS (the survey from which the IALS sample was selected) is about 93 percent, but that it is much lower for certain population subgroups (particularly young Black male adults).

Nonresponse error. For IALS, several procedures were developed to reduce biases due to nonresponse, based on how much of the survey the respondent completed.

Unit nonresponse. The definition of a respondent for IALS was a person who partially or fully completed the background questionnaire. Unweighted response rates varied considerably from country to country, ranging from a high of 69 percent (Canada, Germany) to 45 percent (the Netherlands), with four countries in the 55–60 percent range.

BEST COPY AVAILABLE

In the United States, which had a response rate of 60 percent, nonresponse to IALS occurred for two reasons: (1) some individuals did not respond to the CPS; and (2) some of the CPS respondents selected for IALS did not respond to IALS instruments. In any given month, nonresponse to the CPS is typically quite low, around 4 to 5 percent. Its magnitude in the expiring rotation groups employed for IALS selection is not known. About half of the CPS nonresponse is caused by refusals to participate, while the remainder is caused by temporary absences, other failures to contact, inability of persons contacted to respond, and unavailability for other reasons.

A sizeable proportion of the nonresponse to the IALS background questionnaire was attributable to persons who had moved. For budgetary reasons, it was decided that persons who were not living at the CPS addresses at the time of IALS interviews would not be contacted. This decision had a notable effect on the sample of students, who are sampled in dormitories and other housing units in the CPS only if they do not officially reside at their parents' homes. Those who reside at their parents' homes are included in the CPS at that address, but because most of these students were away at college during the IALS interview period (October to November 1994), they could not respond to IALS.

The high level of nonresponse for college students could cause a downward bias in the literacy skill-level estimates. This group represents only a small proportion of the United States population, however, so the potential bias is likely to be quite small. Further, comparison of IALS results to the U.S. National Adult Literacy Survey data discounts this as a major source of bias.

Item nonresponse. The weighted percentage of omitted responses for the United States IALS ranged from 0 to 18 percent.

Not-reached responses were classified into two groups: nonparticipation immediately or shortly after the background information was collected, and premature withdrawal from the assessment after a few cognitive items were attempted. The first type of not-reached response varied a great deal across countries according to the frames from which the samples were selected. The second type of not-reached response was due to quitting the assessment early, resulting in incomplete cognitive data. Not-reached items were treated as if they provided no information about the respondent's proficiency, so they were not included in the calculation of likelihood functions for individual respondents. Therefore, not-reached responses had no direct impact on the proficiency estimation for subpopulations. The impact of not-reached responses on the proficiency distributions was mediated through the subpopulation weights.

Measurement error. Assessment tasks were selected to ensure that, among population subgroups, each literacy domain (prose, document, and quantitative) was well covered in terms of difficulty, stimuli type, and content domain. The IALS item pool was developed collectively by participating countries. Items were subjected to a detailed expert analysis at ETS and vetted by participating countries to ensure that the items were culturally appropriate and broadly representative of the population being tested. For each country, experts who were fluent in both English and the language of the test reviewed the items and identified ones that had been improperly adapted. Countries were asked to correct problems detected during this review process. To ensure that all of the final survey items had a high probability of functioning well, and to familiarize participants with the unusual operational requirements involved in data collection, each country was required to conduct a pilot survey. Although the pilot surveys were small and typically were not based strictly on probability samples, the information they generated enabled ETS to reject items, to suggest modifications to a few items, and to choose good items for the final assessment. ETS's analysis of the pilot survey data and recommendations for final test design were presented to and approved by participating countries.

Data Comparability

While most countries closely followed the data collection guidelines provided, some did deviate from the instructions. First, two countries (Sweden and Germany) offered participation incentives to individuals sampled for their survey. The incentive paid was trivial, however, and it is unlikely that this practice distorted the data. Second, the doorstep introduction provided to respondents differed somewhat from country to country. Three countries (Germany, Switzerland, and Poland) presented the literacy test booklets as a review of the quality of published documents rather than as an assessment of the respondent's literacy skills. A review of these practices suggested that they were intended to reduce response bias and were warranted by cultural differences in respondents' attitudes toward being tested. Third, there were differences across the countries in the way in which interviewers were paid. No guidelines were provided on this subject, and the study teams therefore decided what would work best in their respective countries. Fourth, several countries adopted field procedures that undermined the objective of obtaining completed background questionnaires for an overwhelming majority of selected respondents.

This project was designed to produce data comparable across cultures and languages. After one of the countries in the first round raised concerns about the international comparability of the survey data, Statistics Canada decided that the IALS methodology should be subjected to an external evaluation. In the judgment of the expert reviewers, the considerable efforts that were made to develop standardized survey instruments for the different nations and languages were successful, and the data obtained from them should be broadly comparable.

However, the standardization of procedures with regard to other aspects of survey methodology was not achieved to the extent desired, resulting in several weaknesses. Nonresponse proved to be a particular weakness, with generally very high nonresponse rates and variation in nonresponse adjustment procedures across countries. For some countries the sample design was problematic, resulting in some unknown biases. The data collection and its supervision differed between participating countries, and some clear weaknesses were evident for some countries. The reviewers felt that the variation in survey execution across countries was so large that they recommended against publication of comparisons of overall national literacy levels. They did, however, despite the methodological weaknesses, recommend that the survey results be published. They felt that the instruments developed for measuring adult literacy constituted an important advance, and the results obtained for the instruments in the first round of IALS were a valuable contribution to the field. They recommended that the survey report focus on analyses of the correlates of literacy (e.g., education, occupation, and age) and the comparison of these correlates across countries. Although these analyses might also be distorted by methodological problems, they believed that the analyses were likely to be less affected by these problems than were the overall literacy levels.

6. CONTACT INFORMATION

For content information on IALS, contact:

Eugene Owen
Phone: (202) 502–7422
E-mail: eugene.owen@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey, NCES 98–053, T.S. Murray, I.S. Kirsch, and L.B. Jenkins (eds.). Washington, DC: 1997.

Chapter 25: National Household Education Surveys (NHES) Program

1. OVERVIEW

The National Household Education Surveys (NHES) Program conducts telephone surveys of the noninstitutionalized, civilian population of the United States. These surveys are designed to provide information on educational issues that are best addressed by contacting households rather than schools or other educational institutions. They offer policymakers, researchers, and educators a variety of statistics on the condition of education in the United States.

Purpose

To (1) provide reliable estimates of the U.S. population regarding specific educational topics, and (2) conduct repeated measurements of the same educational phenomena at different points in time.

Components

The NHES program for a given year typically consists of (1) a screener, which collects household composition and demographic data, and (2) two or three surveys, which are each extended interviews addressing specific education-related topics. However, in 1999, the interviews collected information on key indicators from the broad range of topics addressed in previous NHES survey cycles.

Adult Education and Lifelong Learning. Surveys on this topic were administrated in 2001, 1999, 1995, and 1991.

The Adult Education and Lifelong Learning Survey (AELL-NHES:2001) was administered in 2001. It collected data such as type of program, employer support, and credential sought were collected for participation in the following types of adult educational activities: English as a second language, adult basic education, credential programs, apprenticeships, work-related courses, and personal interest courses. Some information on informal learning activities at work was gathered as well.

In 1999, the Adult Education Survey (AE-NHES:1999) included questions on educational background and work experience, participation in adult education, including educational activities through distance learning, literacy activities, community involvement, adult demographic characteristics, and household characteristics. Eligible respondents were 16 years of age or older who were not currently enrolled in 12th grade or below and not institutionalized or on active duty in the U.S. Armed Forces.

AE-NHES:1995 included questions concerning respondents' participation in basic skills courses, English as a second language (ESL) courses, credential (degree or diploma) programs, apprenticeships, work-related courses, personal development/interest courses,

BIENNIAL SAMPLE SURVEY OF HOUSEHOLD MEMBERS

NHES addresses topical issues on a rotating basis:

- Adult education and lifelong learning
- Before- and after-school programs and activities
- Civic involvement
- Early childhood education and school readiness
- Household library use
- Parent/family involvement in education
- School safety and discipline

and interactive video or computer training on the job. Information on programs or courses included the subject matter, duration, cost, location and sponsorship, and employer support. Nonparticipants in selected types of adult education were asked about their interest in educational activities and barriers to participation. Extensive background, employment, and household information was collected for each adult. Eligible respondents included civilians aged 16 and older not currently enrolled in secondary school.

In AE-NHES:1991, eligible respondents were persons 16 years of age or older, identified as having participated in an adult education activity in the previous 12 months. The information collected on programs and up to four courses included the subject matter, duration, sponsorship, purpose, and cost. A smaller sample of nonparticipants in adult education also completed interviews about barriers to participation. Information on the household and the adult's background and current employment was also collected in this survey.

Before- and After-School Programs and Activities. This survey topic was introduced in 2001. The Before- and After-School Programs and Activities Survey (ASPA-NHES:2001) addressed relative and nonrelative care during the out-of-school hours of school-age children, as well as participation in before- and/or after-school programs, activities, and self-care.

Civic Involvement. Civic involvement surveys were administered in 1999 and 1996. The 1999 Youth Survey (Youth-NHES:1999) expanded on the 1996 Youth Civic Involvement Survey (YCI-NHES:1996). It included questions on school learning environment, family learning environment, plans for future education, participation in activities that promote or indicate personal responsibility, participation in community service or volunteer activities, exposure to information about politics and national issues, political attitudes and knowledge, skills related to civic participation, and type and purpose of community service. A subset of youth who reported participation in community service were asked additional questions about their service experiences. Eligible respondents were youth in the 6th through 12th grades.

Three Civic Involvement Surveys were conducted in 1996: the Parent and Family Involvement in Education/Civic Involvement Survey (PFI/CI-NHES:1996), YCI-NHES:1996, and the Adult Civic Involvement Survey (ACI-NHES:1996). They included questions on sources of political information, civic participation, and knowledge and attitudes about government. YCI-NHES:1996 also provided an assessment of the opportunities that youth have to develop the personal responsibility and skills that would facilitate their taking an active role in civic life. Eligible respondents were (1) parents of students in grades 6 through 12 (including homeschooled students in those grades), (2) youth in grades 6 through 12, and (3) adults.

Early Childhood Education and School Readiness. Early Childhood Education surveys were conducted in 2001, 1995, and 1991, and a School Readiness survey was conducted in 1993.

The Early Childhood Program Participation Survey (ECPP-NHES:2001) was administered in 2001. It gathered information on the nonparental care arrangements and educational programs of preschool children, comprising care by relatives, care by persons to whom they were not related, and participation in day care centers and preschool programs including Head Start.

ECPP-NHES:1995 included questions on children's participation in care or education provided by relatives, nonrelatives, Head Start programs, and center-based programs. It also collected information on early school experiences of school-age children, home literacy activities, health and disability status, and parent and family characteristics. Eligible respondents to this survey were parents of children between birth and 3rd grade. The interview was conducted with the parent most knowledgeable about the child's education or care.

The Early Childhood Education Survey (ECE-NHES:1991) included questions on participation in nonparental care/education, characteristics of programs and care arrangements, and early school experiences including delayed kindergarten entry and retention in grade. In addition to questions about care/education arrangements and school, parents were asked about activities children engaged in with parents and other family members inside and outside the home. Information on family, household, and child characteristics was also collected. Eligible respondents for this survey were the parents or guardians of the sampled 3- to 8-year-olds who were most knowledgeable about the children's education.

The School Readiness Survey (SR-NHES:1993) included questions on the developmental characteristics of preschoolers, school adjustment and teacher feedback to parents for kindergartners and primary school students, center-based program participation, early school experiences, home activities with family members, and health status. Extensive family and child background characteristics—including parents' language and education, income, receipt of public assistance, and household composition—were collected to permit the identification of at-risk children. Eligible respondents to this survey were the parents or guardians of sampled children aged 3 through 7 or in 2nd grade or below who were most knowledgeable about the children's education.

Household Library Use. The Household and Library Use Survey (HHL-NHES:1996) was part of the 1996 NHES screener and consisted of a brief set of questions regarding public library use. Questions addressed the distance to the closest public library, household use of a public library in the past month and year, ways in which the public library was used, purposes for which the public library was used, and detailed household characteristics. Eligible respondents were those adults who completed the Screener interview.

Parent and Family Involvement in Education. Surveys on this topic were conducted in 1996 and 1999. In 1999, the Parent Survey (Parent-NHES:1999) had six sets of questions, appropriate for six subgroups of children: children age 2 and younger, children age 3 through 6 years and not yet in kindergarten, children in kindergarten through the 5th grade, youth in the 6th through 8th grades, youth in the 9th through 12th grades, and children age 5 through 12th grade who were receiving home schooling. The survey included questions on the following topics, although not all topics were covered for all populations: demographic characteristics, current school- or center-based program enrollment status, center-based program participation before school entry, home schooling, school characteristics, school readiness skills, participation in early childhood care and programs, training and support for families of preschoolers, parents' satisfaction with children's schools, children's academic performance and behavior, family involvement with children's schools and school practices to involve families, before- and after-school programs and nonparental care, parents' expectations about children's college plans and costs, family involvement in educational activities outside of school, child health and disability, parent/guardian characteristics, and household characteristics. The Parent Survey was administered to the parent or guardian most knowledgeable about the education of each sampled child from birth through 12th grade.

In 1996, the survey was combined with one on Civic Involvement, forming PFI/CI-NHES:1996. It included questions on the schools of the sampled children, communication with teachers or other school personnel, school practices to involve parents, children's homework and behavior, and learning activities with children outside of school with their families. Other information collected in this survey pertain to student experiences in school, children's personal and demographic characteristics, household characteristics, and children's health and disability statuses. Eligible respondents were the parents or guardians of children aged 3 through 20 and in 12th grade or below who were most knowledgeable about the child's education.

School Safety and Discipline. The School Safety and Discipline Survey (SS&D-NHES:1993) included questions on school learning environment, discipline policy, safety at school, victimization, availability and use of alcohol/drugs, and alcohol/drug education. Peer norms for behavior in school and substance use were also included in this survey. Extensive family and household background information and data about characteristics of the school attended by the child were collected. Eligible respondents were the parents or guardians of sampled children in grades 3 through 12 and youth in grades 6 through 12 who were most knowledgeable about the child's education.

Periodicity

Biennial as of 1999. Earlier surveys were conducted in 1991, 1993, 1995, and 1996.

2. USES OF DATA

NHES provides descriptive data on the educational activities of the U.S. population and offers policymakers, researchers, and educators a variety of statistics on the condition of education in the United States. Each NHES survey collects specific data based on a set of research questions that guide the development of the questionnaire. As described above, the main subject areas for the NHES programs are:

- Adult Education and Lifelong Learning
- Before- and After-School Programs and Activities
- Civic Involvement
- Early Childhood Education and School Readiness
- Household Library Use
- Parent and Family Involvement in Education
- School Safety and Discipline

Analysts should review the instrument for each survey to identify areas of particular interest to them.

3. KEY CONCEPTS

See the survey documentation for definitions specific to any one NHES survey.

Household Members. Individuals who think of the sampled household as their primary place of residence, including persons who usually stay in the household but are temporarily away on business or vacation, in a hospital, or living at school in a dormitory, fraternity, or sorority.

4. SURVEY DESIGN

Target Population

Noninstitutionalized, civilian members of households in the 50 states and the District of Columbia. Because the topical surveys change from one NHES to the next, the specific age/grade criteria for the target populations also change. In general, there are three educational populations of interest: (1) younger children from birth through 5th grade; (2) older children (i.e., youth) in the 6th through 12th grades; and (3) adults not enrolled in 12th grade or below. The respondent is usually the parent or guardian of the child who is most knowledgeable about the education or care of the sampled child, the sampled youth, or the sampled adult.

Sample Design

The NHES samples are selected using random-digit-dialing (RDD) methods. Telephone numbers are randomly sampled, and a screener is administered to sampled households. About 45,000 to 64,000 households are screened for each administration. Individuals within households who meet predetermined criteria are then sampled for more detailed or extended interviews.

Sampling households. Two general sampling approaches have been taken: list assisted and a modified Mitofsky-Waksberg method. The list-assisted method has been used since the 1995 administration.

In 2001, a two-phase list-assisted method was used. In the first phase of selection, telephone numbers were stratified according to the percent minority in the exchange. Exchanges with at least 20 percent Blacks or at least 20 percent Hispanics were classified as "high minority" and all other exchanges were classified as "low minority." Telephone numbers in the high minority stratum were sampled at a rate of about 1 in 809, and telephone numbers in the low minority stratum were sampled at a rate of about 1 in 1,562. The first phase sample of telephone numbers was processed using the Genesys ID-Plus process to identify nonworking and business numbers. As part of this process, the telephone numbers were matched to white pages listings, and the matches were flagged. Thus, for each telephone number in the first phase sample, the listed status (i.e., whether or not it is listed in the white pages) is known. Within each minority stratum, the telephone numbers in the first phase sample were stratified according to white pages listed status (the overall number of telephone numbers selected in phase 1 was 206, 182). At the second phase, telephone numbers within each of the four strata defined by the combinations of minority concentration and listed status were subsampled at different rates: 0.714 for the high minority, listed stratum; 0.950 for the high minority, unlisted stratum; 0.727 for the low minority, listed stratum; and 0.942 for the low minority, unlisted stratum. The total number of telephone numbers selected in phase 2 was 179, 211.

A list-assisted method was used in the 1995, 1996, and 1999 administrations. This approach involves selecting a simple random sample of telephone numbers from all telephone numbers in 100-banks (i.e., sets of numbers with the same first 8 digits of the 10-digit telephone number) that have at least one telephone number listed in the white pages (called the listed stratum). Telephone numbers in 100-banks with no listed telephone numbers (called the zero-listed stratum) are not sampled. Because the list-assisted approach is an unclustered design, it results in estimates with lower variances than the clustered alternative methods. However, this method also incurs a small amount of coverage bias because households in the zero-listed stratum have no chance of being included in the sample. (See section 5, "Coverage error" for a discussion of coverage bias. See "Stratified Telephone Survey Designs," by R.J. Casady and J.M. Lepkowski, in *Survey Methodology 19*(1) (1993): 103–113, for further description of the list-assisted method.)

For the surveys fielded in 1996, the goal of making estimates at the state level for characteristics of household members and for household library use also determined the number of telephone numbers selected. A target of 500 screened households per state was set. A sample of 500 households is large enough that, if 30 percent of the households in a state have a given characteristic, differences of 6 percent can be detected. Due to nonresponse at the screener level and lower residency rates than expected in some states, 500 screeners were not completed in some states. The lower number of responses limits the ability to make estimates for some subgroups within states. Analysts should examine the standard errors for subgroups of interest to evaluate the precision of within-state estimates.

The NHES surveys fielded in 1991 and 1993 used a modified version of the Mitofsky-Waksberg method of RDD, in which a fixed number of telephone numbers is sampled from 100-banks. (See "Avoiding Sequential Sampling with Random Digit Dialing" by J.M. Brick and J. Waksberg, *Survey Methodology 17*(1) (1991): 27–42 for further description of the modified Mitofsky-Waksberg method used in the NHES.)

Oversampling households for Blacks and Hispanics. One of the goals of the NHES program is to produce reliable estimates for subdomains defined by race and ethnicity. In a 64,000-household design in which every household has the same probability of being included, the number of completed interviews would not be large enough to produce reliable estimates of many characteristics of Black and Hispanic youth. Therefore, in each NHES administration, telephone numbers in areas with high concentrations of Blacks and Hispanics are oversampled. In 1993, areas with high percentages of Asians were also sampled at a higher rate; this was discontinued in later administrations because the new vendor for numbers on the list-assisted approach of sampling did not have this information available. NHES considered reintroducing an Asian oversampling strategy in 2001. However, it was determined that more precision in other racial/ethnic groups would have been lost than was warranted given the amount of extra precision gained for Asians.

A computer file containing census characteristics for telephone exchanges is used to stratify telephone exchanges into low- and high-minority concentration strata. Any telephone exchange not found on the file is assigned to the low-minority concentration stratum. High-minority concentration areas are defined as exchanges having at least 20 percent Black or 20 percent Hispanic persons living in the area (or 20 percent Asian/Pacific Islander persons for the 1993 NHES). The telephone exchanges in the two strata are identified, and a systematic sample is drawn in each stratum. The sampling fraction used in the high-minority concentration stratum is two times the fraction used in the low-minority concentration stratum.

Oversampling by the characteristics of the telephone exchange has two effects. First, the oversampling increases the sample sizes for minorities because they are more heavily concentrated in the exchanges that are oversampled. Therefore, the sampling errors for estimates of these groups are reduced due to the increased sample sizes. On the other hand, not all minorities are found in the oversampled exchanges. Thus, differential sampling rates are applied to persons depending on their exchanges. Using differential rates increases the sampling errors of the estimates, partially offsetting the benefit of the larger minority sample. However, the net result is an increase in precision of estimates for Black and Hispanic persons. The technical report *Effectiveness of Oversampling Blacks and Hispanics in the NHES Field Test* (NCES 92–104) indicates that oversampling is successful in reducing the variances for estimates of characteristics of Blacks and Hispanics by approximately 20 to 30 percent over a range of statistics examined. The decreases in precision for estimates of the groups that are not oversampled and for estimates of totals are modest, ranging from about 5 to 15 percent.

Approaches to household enumeration. The approach to screening households has also changed over the course of the NHES program. Changes include methods of enumerating members of households that are contacted and the amount of information collected in the screener about the household and its members. In 1991, a split-enumeration design was used; all households were screened for ECE-NHES:1991, and a subset of households was screened for AE-NHES:1991. In 1993, when SR-NHES:1993 and SS&D-NHES:1993 were fielded, households were enumerated only when there were any household members aged 20 or younger. The only information collected in both 1991 and 1993 was the first name, age, and sex of household members. In both 1995 and 1996, *all* screened households were fully enumerated. The 1995 administration included a test of an expanded screener that was used in 1996, but dropped from later NHES administrations. The 1996 screener collected educational and demographic information on household members and included a brief topical survey. The 1999 screener again collected first name, age, and sex of household members, but not all households were fully enumerated, so if the screener respondent said there were no children in the household and the household was not preselected to be eligible for an adult education interview, the screener information was not collected.

Sampling within households. The within-household sample designs for the NHES collections are determined

BEST COPY AVAILABLE

by the specific goals of the surveys administered and by the combination of surveys administered in a specific year. Brief summaries of the within-household sampling for the various NHES administrations are given below, by year.

2001 NHES surveys—AELL-NHES:2001, ASPA-NHES:2001, and ECPP-NHES:2001. A within-household sample scheme was developed to control the number of persons sampled for extended interviews in each household. The sample of telephone numbers was randomly divided into three groups. The first group (89,597 telephone numbers or approximately 50 percent of the sample) was designated for adult enumeration. The second group (44,985 telephone numbers or about 25 percent of the sample) was designated for adult enumeration only if there were no eligible children in the household. The third group (44,629 telephone numbers or about 25 percent of the sample) was designated for no adult enumeration. Once the enumeration of the appropriate household members was completed in the Screener, the sample of household members for the extended interviews was done by computer. The ECPP and ASPA interviews were conducted with parents/guardians of sampled children from birth through age 15 who were in 8th grade or below. In households with one or more preschoolers (children age 3 through 6 and not yet in kindergarten), one child in this age/grade range was sampled. In households with middle school students (6th through 8th grades), one child in this age/grade range was also sampled. The sampling of infants (newborn through age 2), elementary school children (kindergarten through grade 5), and adults was conducted using an algorithm designed to attain the sample rates required to meet the target sample sizes while minimizing the number of interviews per household. The within-household sample size was limited to three eligible children if no adults were to be selected or two eligible children and one eligible adult. No more than one child from any given domain (i.e., infants, preschoolers, elementary students, middle school students) was sampled in any given household. This sampling algorithm was designed to limit the amount of time required to conduct interviews with parents in households with a large number of eligible children.

1999 NHES surveys—AE-NHES:1999, Parent-NHES:1999, and Youth-NHES:1999. The overall screening sample was largely determined by the need to produce precise estimates of indicators for young children, particularly preschoolers. Since sample requirements were most stringent for preschoolers (children ages 3–6 not yet in kindergarten), it was decided to sample one preschooler in every household that had such children. Another goal was that no more than three persons per household be sampled, with a maximum of four extended interviews per household. To accomplish this, several flags were set prior to screening. The first specified whether adults in the household were to be enumerated, as well as the conditions under which an adult was to be sampled. This flag was set such that households without eligible children/youth were sampled for an Adult Education Survey at approximately twice the rate of households with eligible children/youth (about 26 percent vs. 13 percent). Additionally, this flag enabled one- and two-adult households with no adult education participants to be further subsampled at a fixed, prespecified rate (25 percent for one-adult households and 75 percent for two-adult households). The second flag designated whether an infant was to have been sampled, if the household had two other sampled children/youth. A third flag designated whether a younger child or an older child was to be sampled, if the household had children in both groups, only one was to be selected. In households in which an adult was to be sampled, each adult education participant was given a probability of selection 2.5 times as large as the probability of selection assigned to nonparticipants.

1996 NHES surveys—ACI-NHES:1996, HHL-NHES:1996, PFI/CI-NHES:1996, and YCI-NHES:1996. The number of interviews for which household members could be selected was limited by creating two separate samples—Parent/Youth and Adult. A sample of 161,446 telephone numbers was selected and randomly divided into two groups. The first group (153,374 telephone numbers or 95 percent of the sample) was allocated to the Parent/Youth sample. A screening interview was conducted in these households, and eligible children and youth were sampled, respectively, for PFI/CI-NHES:1996 or for both PFI/CI-NHES:1996 and YCI-NHES:1996. For PFI/CI-NHES:1996, if there were one or more children from age 3 through 5th grade (younger children), one child in this age range in the household was sampled for the survey. If the household included one or more children in 6th through 12th grades (older children), one child in this grade range was sampled from that household. If an older child was sampled as the subject of a PFI/CI-NHES:1996 interview, the child was asked to complete the YCI-NHES:1996. Because households may have had up to two Parent PFI/CI interviews (one for a younger child and one for an older child), the maximum number of interviews per sampled household was three. The other group (8,072 telephone numbers or 5 percent of the

BEST COPY AVAILABLE

sample) contained those telephone numbers allocated to the ACI-NHES:1996. For households in that group, a screening interview was conducted and the ACI-NHES:1996 was administered to one eligible adult.

1995 NHES surveys—AE-NHES:1995 and ECPP-NHES:1995. Interviews for ECPP-NHES:1995 were conducted with the parents or guardians who were most knowledgeable about the education of the sampled children aged 0 to 10 years who were in the 3rd grade or below. The within-household sample size was limited to two eligible children. Children in kindergarten were sampled at 1.5 times the rate for other children to improve the precision of single-year estimates for kindergartners. Any adult aged 16 years or older not currently enrolled in secondary school was eligible for sampling for AE-NHES:1996. Sampled adults who said they were on active duty in the U.S. Armed Forces were classified as ineligible for the interview.

1993 NHES surveys—SR-NHES:1993 and SS&D-NHES:1993. For the 1993 NHES surveys, children within households were subsampled. For SR-NHES:1993, interviews were conducted with the parents or guardians who were most knowledgeable about the education of children aged 3 through 7 and children aged 8 or 9 who had not completed 2nd grade. If there were one or two eligible children in a household, all the children were sampled. If there were more than two, two were randomly sampled from the household. Any child enrolled in grades 3 through 12 and below the age of 21 was eligible for sampling for the SS&D-NHES:1993 interview with the parent. Sampling was limited to one child in 3rd through 5th grades and no more than two children in any household. No more than one youth was subsampled per household for the youth interview. If a child was enrolled in the 6th through 12th grades but did not live with a parent or guardian, he or she was considered an emancipated youth. A special emancipated youth interview was conducted, including some questions usually asked only of parents.

1991 NHES surveys—AE-NHES:1991 and ECE-NHES:1991. All 3- to 8-year-olds in sampled households were included in ECE-NHES:1991, as were 9-year-olds who had not completed 2nd grade. All children 2 to 9 years old were sampled to ensure that nearly all children eligible for the extended interviews were identified, even if a rounding error was made in reporting the ages of the children. The respondent for the interview was the parent or guardian of the sampled child reported to be the most knowledgeable about the child's care and education. Only a subset of households was screened for AE-NHES:1991. In the screened households, all adults identified as participating in adult education activities were sampled, half of the full-time degree-seeking students were sampled, and about 7 percent of the nonparticipants in adult education activities were sampled. After a few weeks of data collection, the number of sampled households screened for AE-NHES:1991 was reduced because the required number of interviews had been completed and therefore additional households did not need to be contacted; altogether, 18,463 households out of 60,300 completed screeners (31 percent) were screened for AE-NHES:1991. In addition, the sampling rate for nonparticipants was increased from 7 percent to 12 percent.

Data Collection and Processing

NHES program surveys are conducted using computer-assisted telephone interviewing (CATI). Westat has been the contractor on all surveys to date.

Reference dates. Most data items refer to the time of data collection or since September of the current school year. Other items are asked retrospectively for different time frames. For example, in the 1996 NHES surveys, respondents were asked about family involvement with children outside of school (e.g., reading with a child, visiting a library) in the past week and past month; civic involvement (reading about or watching national news) in the past week; political activities in the past 12 months; voting activities in the past 5 years; working for pay during the past week and the past 12 months; job-hunting in the past 4 weeks; child's communications with the noncustodial parent in a typical month and in the past year; youth's discussion of future educational plans with parents in the past month; books read in the past 6 months; home visits by professionals during the past 12 months; and religious service participation in the past year. The adult education information is based on participation in the past 12 months.

Data collection. Data collection for the NHES surveys takes place over a 3- to 4-month period beginning in January of each survey year. The data are collected using CATI. The NHES screeners are completed with an adult household member in households selected using random-digit-dialing techniques. (See Sample Design above.)

Over a period of about 3 weeks just prior to data collection, more than 300 interviewers undergo intensive training in general interviewing techniques, use of the CATI system, and the conduct of the survey.

Most responses were coded at the time of the interview. Most of the items in the surveys are close ended, meaning respondents are given a short list of response options. Interviews simply record the response as a one- or two-digit code which is entered directly into the data file as the interview progresses. However, most close-ended items do have "other, specify" options that allow interviews to record responses that do not fit the precoded response categories. The interviewer types in these "open-ended" responses as one or more sentences. "Other, specify" responses to close-ended items are rare. There are also a small number of items in some of the surveys that are designed to be open ended. That is, precoded categories do not exist and interviewers type in verbatim responses from respondents. Once the survey is completed, data preparation staff and survey managers review these open-ended responses to determine how they can be coded into a limited set of response categories. Coding of additional items was required for the Adult Education surveys administered in 1991 and 1995. These items included adult education courses, major fields of study for college and vocational programs, industry, and occupation. A double-blind coding procedure was used, in which two coders independently assigned a code to the response. When the coding was discrepant, an "adjudication" coder reviewed the case and assigned an appropriate final code.

Editing. Intensive data editing is a feature of both the data collection and file preparation phases of the NHES collections. Range checks for allowable values and logic checks for consistency between items are included in the online CATI interview so that many unlikely values or inconsistent responses can be resolved while the interviewer is speaking with the respondent.

Postinterview editing is conducted throughout data collection and after data collection is completed. In addition to range and logic edits, the postinterview editing process includes checks for the structural integrity of the hierarchical CATI database and integrity edits for complex skip patterns. It also includes a review of comments provided by interviewers and problem sheets completed by interviewers. Following the resolution of any problems, data preparation staff review frequency distributions and crosstabulations of the data sets in order to identify any remaining skip pattern inconsistencies. Editing is repeated following completion of imputation.

Estimation Methods

The NHES surveys use weighting to adjust for the fact that the sampling is not simple random sampling. It is also used to adjust for potential undercoverage bias and potential unit nonresponse bias. Imputation is performed to compensate for item nonresponse.

Weighting. The objective of the NHES surveys is to make inferences about the entire noninstitutionalized, U.S. civilian population and about subgroups of interest. Although only telephone households are sampled, the estimates are adjusted to totals of persons living in both telephone and nontelephone households derived from the Current Population Survey (CPS) to achieve this goal. (CPS is an annual household survey conducted by the U.S. Bureau of the Census for the U.S. Bureau of Labor Statistics.) As a result, any undercoverage in CPS for special populations, such as the homeless, are also reflected in NHES estimates. The potential for bias due to sampling only telephone households has been examined for virtually all the population groups sampled in NHES. Generally, the bias in the estimates due to excluding nontelephone households is small. (See section 5, "Coverage error," for further discussion.) The weighting procedures across NHES surveys are very similar. Weighting consists of two stages: household-level weighting and person-level weighting, as described below.

Household weights. The household weights take into account all factors that might have resulted in adjustments due to the telephone numbers being sampled at different rates. Two factors common to all NHES years are (1) the adjustment to account for the differential sampling rates by minority concentration and (2) the adjustment to account for households that have more than one telephone number and, hence, chance of being sampled. In 1991 and 1993, an adjustment was also made to account for the modified Mitofsky-Waksberg method of random-digit-dialing sampling. (See earlier section on Sample Design.) The 1996 NHES included an adjustment for the oversampling in 18 states to bring the minimum expected number of completed screeners up to 500.

Response rates declined after 1993, requiring analyses to be conducted to study if nonresponse bias was becoming a significant problem in the data. For example, for the 1995 administrations, the variables correlated with the response rate were then used to define nonresponse adjustment classes, and the inverse of the response rate in a class was used as the weight adjustment. The nonresponse adjustment classes were based on the following variables: metropolitan status, census division, percent renters, percent owner occupied, percent college graduates, median income, percent Black, percent Hispanic, and

BEST COPY AVAILABLE

percent aged 0 to 17. The nonresponse-adjusted weights were then used in all other stages of weighting in the 1995 NHES surveys in an effort to reduce nonresponse bias. Similar analyses were conducted in later years.

In 1996, for the first time, household weights were needed to produce estimates from the Household and Library data file. The 1996 household weights were adjusted to known national totals of households using raking to ensure that the estimates conformed to national totals, to reduce the bias associated with sampling only telephone households, and to adjust for nonresponse bias. As a result of raking, the household estimates match control totals of the number of households within each state and the District of Columbia defined by the following dimensions: the presence of children under age 18, owned or rented home, urban or rural location, and race of the oldest household member (not taking into account Hispanic ethnicity). The control totals were the March 1995 CPS total household estimate distributed according to the 1990 decennial Census of Population and household distributions. In some states, all four of the dimensions were defined and used for raking; in other states, only three dimensions were used because the expected number of completed screeners fell below 50 in a given cell when four dimensions were considered. NHES also raked household weights to national totals for the 1999 and 2001 surveys. The approach used was similar to that described above, but the control totals were from the March 1998 CPS and March 2000 CPS, respectively.

Person weights. The second stage of weighting forms person weights for each extended interview. For example, in 1991, person weights were developed for each sampled child in ECE-NHES:1991 even if the same parent responded to both interviews. Thus, the estimates from this survey correspond to the population of children eligible for the survey. Person weights are prepared for each extended interview in every NHES program survey.

The first step in creating the person weights is to assign the appropriate household weight to the sampled person as a base weight that can then be modified to account for other stages of sampling, nonresponse, and adjustments to known population totals from CPS. The first modification to the base weight accounts for the within-household sampling of persons. The appropriate sampling factor for each survey and survey year is multiplied by the base weight to produce an initial person weight for each completed interview.

The second step is to adjust the person weights to account for nonresponse. This step was not necessary for ECE-NHES:1991 and ECPP-NHES:1995 because the completion rates were so high for all the sampled children. In most of the surveys, some characteristics about the sampled person are collected in the screener and used to form nonresponse adjustment classes. These characteristics include age, sex, grade in school, adult education participation status, and education level. The nonresponse adjustment for respondents within a class is the inverse of the within-class completion rate for the extended interviews. If the completion rates for a survey do not vary much from one class to the next, the nonresponse adjustments are relatively constant over the classes. Adjustments can vary substantially if there is greater variation in completion rates. There was a person-level nonresponse adjustment in ECE-NHES:2001.

The third and final step in developing person weights is the raking of the nonresponse-adjusted person weights so that the survey estimates match appropriate control totals for the population being surveyed. This raking procedure is identical to the one described above for the final household weights in the NHES surveys administered in 1996, the only difference being the substitution of person weights and counts for household weights and counts. The source of the control totals for the number of persons is the CPS for the month corresponding most closely to the NHES survey for which comparable estimates can be produced. For the NHES surveys administered in 1996, however, the weights were raked to national totals obtained by multiplying the percentage distributions from the October 1994 CPS (which contained additional variables) by the estimates of the number of children from the March 1995 CPS (the most current population data). Although the variables used to form the control totals vary from year to year and survey to survey, they are very similar since the main purpose of the raking is to reduce the bias in the estimates arising from the failure to sample nontelephone households. Typically, the control totals involve some combination of the following variables: home owned or rented, race/ethnicity, household income, Census region (Northeast, South, Midwest, West), urban or rural location, and age or grade. The final person weights on the public release data files are the raked person weights. The same October CPS/March CPS approach was used in all other collection years as well.

Imputation. Item response rates for most data items collected in NHES surveys are very high. Nevertheless, virtually all items with missing data (including "don't know" and "refused" responses) are imputed in NHES surveys. In the two NHES surveys administered in 1991, only variables that were used for the development of weights or derived variables were fully imputed. Text responses (for example, in Youth-NHES:1999, type of service activity, or, in AE-NHES:1999, name of company) were not imputed in any year. Occasionally, "don't know" and "refused" responses are of analytic interest so are not imputed. For example, in the Youth-NHES:1999 survey, "don't know" and "refused to answer" responses to the knowledge about government items were not imputed.

Imputations are done in the NHES program for three reasons. First, complete responses are needed for the variables used in developing the sampling weights. Second, users compute estimates employing a variety of methods, and complete responses should aid their analysis. Third, imputation may reduce bias due to item nonresponse, by obtaining imputed values from donors that are similar to the recipients. The procedures for imputing missing data are discussed below.

A standard (random within class) hot-deck procedure has been used to impute missing responses in every NHES collection. In this approach, the entire file is sorted into cells defined by characteristics of the respondents. The variables used in the sorting are general descriptors of the interview and also include any variables involved in the skip pattern for the items. All of the observations are sorted into cells defined by the responses to the sort variables, and then divided into two classes within the cell depending on whether or not the item being imputed is missing. For an observation with a missing value, a value from a randomly selected donor (observation in the same cell but with the item completed) is used to replace the missing value. After the imputation is completed, edit programs are run to ensure that the imputed responses do not violate edit rules.

For some items, the missing values are imputed manually rather than using the hot-deck procedure. This happens most often when the variable is collected only once for the household or involves complex relationships. Manual imputation is also used if a small number of edit failures are found after the hot-deck imputations are completed. In the 1999 NHES surveys, manual imputation was done to (1) impute certain person-level characteristics from the screener; (2) impute whether a child is homeschooled, if the child attends regular school for some classes, and the number of hours the child attends regular school; (3) correct for a small number of inconsistent imputed values; (4) impute for a few cases when no donors with matching sort variable values could be found.

After values have been imputed for all observations with missing values, the distribution of the item prior to imputation (i.e., the respondent's distribution) is compared to the post-imputation distribution of the imputed values alone and of the imputed values together with the observed values. This comparison is an important step in assessing the potential impact of item nonresponse bias and ensuring that the imputation procedure reduces this bias, particularly for items with relatively low response rates (less than 90 percent).

For each data item for which any values are imputed, an imputation flag variable is created so that users can identify imputed values. Users can employ the imputation flag to delete the imputed values, use alternative imputation procedures, or account for the imputation in computation of the reliability of the estimates produced from the data set.

Recent Changes

A two-phase sample design was used in the NHES surveys administered in 2001, and the NHES program adopted a new procedure for replication variance estimation for two-phase samples.

Future Plans

According to the current plan, three surveys will be included in each NHES collection. In 2001, and at subsequent 4-year intervals, the surveys will be Early Childhood Program Participation, Before- and After-School Programs and Activities, and Adult Education and Lifelong Learning. In alternate collections (2003 and subsequent 4-year intervals), the surveys will be School Readiness, Parent and Family Involvement in Education, and Adult Education for Work-Related Reasons. However, in 2003, School Readiness will not be fielded due to budgetary constraints.

5. DATA QUALITY AND COMPARABILITY

In addition to the data quality activities inherent in the NHES design and survey procedures, activities specifically designed to assess the quality of data are undertaken for each collection. Reinterviews and analysis of telephone coverage bias are two activities conducted during every survey administration. Other data quality activities address specific concerns related to a topical survey. Issues of data quality and comparability are discussed below.

Sampling Error

The two major methods of producing approximate standard errors for complex samples are replication methods and Taylor Series approximations. Special software is available for both methods, and the NHES data support either type of analysis. (Further information on the use of replication and Taylor Series methods is provided in *A Guide to Using Data from the National Household Education Survey (NHES)*, NCES 97–561.)

Since the 2001 NHES surveys used a two-phase sample design, a new procedure for replication variance estimation was also used. The replicate base weights under two-phase sampling are calculated using a two-step procedure. First, the initial replicate base weights of the first-phase units are calculated using the standard jackknife procedure. In the second step, the final replicate base weights for the second-phase sample are computed by redistributing the initial replicate weights of first-phase units not selected in the second phase to the initial replicate weights of the second-phase units within the same second-phase stratum. That is, for unit *i*, the replicate weight for the *j*th replicate is

$$w_{2i}^{j} = \begin{cases} w_{1i}^{j} \dfrac{\sum\limits_{k \in S_{1h}} w_{1k}^{j}}{\sum\limits_{k \in S_{2h}} w_{1k}^{j}}, & i \in S_{2h} \\ \\ 0, & i \notin S_{2h} \end{cases}$$

where

h denotes the second phase stratum,

S_{1h} denotes the first phase sample in stratum h,

S_{2h} denotes the second phase sample in stratum h, and

w_{1i}^{j} denotes the initial replicate j base weight for unit i.

Note that the sum of the final replicate base weights of the second-phase units is the same as the sum of the initial replicate base weights of the first-phase units within the same second-phase stratum. The procedure involves only the calculation of the telephone number-level replicate base weights. All full-sample weighting and all subsequent adjustments to the replicate weights are done using the same methodology used for a single-phase sample.

The *replication method* used in the NHES surveys for single-phase samples involves splitting the entire sample into a set of groups, or replicates, based on the actual sample design of the survey. The survey estimates can then be estimated for each of the replicates by creating replicate weights that mimic the actual sample design and estimation procedures used in the full sample. The variation in the estimates computed from the replicate weights can then be used to estimate the sampling errors of the estimates from the full sample. The procedures used to develop the full weights are used to produce each replicate weight. Replicate weights have been included in all of the NHES data files to make this application relatively simple. Various software packages such as WesVar, SUDAAN, etc. can properly apply replicate weights.

Nonsampling Error

Sample estimates also are subject to bias from nonsampling errors. It is more difficult to measure the magnitude of these errors. They can arise for a variety of reasons: nonresponse; undercoverage; differences in the respondent's interpretation of the meaning of questions; memory effects; misrecording of responses; incorrect editing, coding, and data entry; time effects; or errors in data processing.

Coverage error. Every household survey is subject to some undercoverage bias—the result of some members of the target population being either deliberately or inadvertently missed in the survey. Telephone surveys like those in the NHES program are subject to an additional source of bias because not all households in the United States have telephones: approximately 6 percent of adults aged 16 years or older (and not enrolled in elementary or secondary school) and about 7 percent of children age 20 or younger and in grade 12 or below live in households without telephones. Even more problematic is the fact that the percentage of households without telephones varies from one subgroup of the population to another. If all telephone households are included in the survey and respond to the required interviews, the difference between the estimate from the survey and the actual population value (which includes the responses of persons living in nontelephone households) is the bias due to incomplete coverage. Since NHES surveys are based on a sample, the bias is defined as the expected or average value of this difference over all possible samples.

Special analyses of the bias associated with telephone coverage and its potential impact on estimates from the NHES surveys are conducted for each cycle of the

BEST COPY AVAILABLE

survey. Data from CPS are used to evaluate the differences between estimates for telephone households and estimates for the entire population. (CPS is an annual household survey conducted by the U.S. Bureau of the Census for the U.S. Bureau of Labor Statistics.) The results of these analyses show that, for most estimates, the bias due to sampling only telephone households is small. However, for subgroups with characteristics highly correlated with not having a telephone (e.g., the poor, high school dropouts), coverage bias may be large. Raking adjustments do often reduce such coverage bias, though no adjustments have been found to adequately reduce the amount of bias across all measures that might be affected by coverage issues. (See, for example, *Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey,* NCES 97–39.)

Additional undercoverage results when some telephone households are excluded from the sampling frame. This was a disadvantage of the list-assisted method of random-digit-dialing sampling used in earlier administrations of NHES surveys. (See section 4, Sample Design.) Households in the zero-listed stratum had no chance of being included in the sample. Empirical findings that address questions of coverage bias show that the percentage of telephone numbers in the zero-listed stratum that are residential is very small (about 1.4 percent) and that about 3 to 4 percent of all telephone households are in the zero-listed stratum. The findings also show that the bias resulting from excluding the zero-listed stratum is generally small. (See "Bias in List-assisted Telephone Samples," by J. M. Brick, J. Waksberg, D. Kulp, and A. Starer, in *Public Opinion Quarterly 59*(2) (1995): 218–235.)

Nonresponse error. Nonresponse in NHES surveys is handled in ways designed to minimize the impact on data quality—through weighting adjustments for unit nonresponse and through imputation for item nonresponse.

Unit nonresponse. Household members are identified for extended interviews in a two-stage process. First, screener interviews are conducted to enumerate and sample households for the extended interviews. The failure to complete the first-stage screener means that it is not possible to enumerate and interview members of the household. The completion rate for the first stage is the percentage of screeners completed by households. The completion rate for the second stage is the percentage of sampled and eligible persons with completed interviews. The survey response rate is the product of the first- and second-stage completion rates (*screener completion rate* x *interview completion rate = survey response rate*, see table 11, on the next page). All of the rates are weighted by the inverse of the probability of selecting the units.

Item nonresponse. For most of the items collected in the NHES surveys, the item response rate is high. The median item response rate for items with any missing values for the surveys administered in 1995, 1996, and 1999 ranged from 98.4 to 99.5, except for HHL-NHES:1996, where the median response rates for imputed items was 95.0 for household-level characteristics and 99.5 for person-level characteristics. For SR-NHES:1993, three items had response rates of less than 95 percent; for SS&D-NHES:1993, there were two such items. None of the ECE-NHES:1991 items had response rates of less than 94 percent, while most of the AE-NHES:1991 items had response rates of more than 99 percent; however, there was one item from the 1991 screen which had a response rate of 92 percent.

Measurement error. In order to assess item reliability and inform future NHES surveys, most administrations also include a subsample of respondents for a reinterview. Reinterviews were conducted for ECE-NHES:1991, both SR-NHES:1993 and SS&D-NHES:1993, AE- NHES:1995, and both Parent-NHES:1996 and Youth-NHES:1996.

In a reinterview, the respondent is asked to respond to the same items on different occasions. In order to limit the response burden of the reinterview program, only selected items are included in the reinterview. The item selection criteria focus on the inclusion of key survey statistics (e.g., frequency of reading to children), items that are expected to have a potential for measurement error based on cognitive laboratory or field test findings, and items required to control the question skip patterns for the reinterview. The results of the reinterviews are used to modify subsequent NHES surveys and to give some guidance to users about the reliability of responses for specific items in the data files. (See *Use of Cognitive Laboratories and Recorded Interviews in the National Household Education Survey,* NCES 96–332.) However, the reinterview procedure does not account for all measurement errors in the interviewing process, such as systematic errors that would be made in both the original interview and the reinterview.

The major emphasis of the 1991, 1993, and 1995 reinterview studies was to measure response variability. Overall, the results were positive. For example, within the AE-NHES:1995 reinterview study, only three items

Table 11. Weighted response rates for selected NHES surveys

Questionnaire	Screener/1st stage	Interview/2nd stage	Overall
ECE-NHES:1991	81.0	94.5	76.5
AE-NHES:1991	81.0	84.7	68.6
SR-NHES:1993	82.1	89.6	73.6
SS&D-NHES:1993 – Parents, 3rd–5th	82.1	89.4	73.4
SS&D-NHES:1993 – Parents, 6th–12th	82.1	89.6	73.6
SS&D-NHES:1993 – Students, 6th–12th	82.1	83.0	68.1
ECPP-NHES:1995	73.3	90.4	66.3
AE-NHES:1995	73.3	80.0	58.6
PFI/CI-NHES:1996	69.9	89.4	62.5
YCI-NHES:1996	69.9	76.4	53.4
ACI-NHES:1996	69.9	84.1	58.9
Parent-NHES:1999	74.1	90.0	66.7
Youth-NHES:1999	74.1	78.1	57.9
AE-NHES:1999	74.1	84.1	62.3
AELL-NHES:2001	69.2	77.2	53.4
ECPP-NHES:2001	69.2	86.6	59.9
ASPA-NHES:2001	69.2	86.4	59.7

SOURCE: Brick and Broene, *Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)* (NCES Working Paper 97–06). Brick, Collins, Celebuski, Nolin, Squadere, Ha, Wernimont, West, Chandler, Hausken, and Owings, *National Household Education Survey Adult and Course Data Files User's Manual* (NCES 92–019). Brick, Collins, Celebuski, Nolin, Squadere, Ha, Wernimont, West, Chandler, Hausken, and Owings, *National Household Education Survey Preprimary and Primary Data Files User's Manual* (NCES 92–057). Brick, Tubbs, Collins, and Nolin, *Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)* (NCES Working Paper 97–05). Collins, Montaquila, Nolin, Kim, Kleiner, and Waits, *National Household Education Surveys of 2001 Data File User's Manual, Volume I* (forthcoming). Montaquila and Brick, *Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey* (NCES Working Paper 97–40). Nolin, Montaquila, Nicchitta, Kim, Kleiner, Lennon, Chapman, Creighton, and Bielick, *NHES:1999 Methodology Report* (NCES 2000–078).

in one subject area had high response variability. The reinterview responses were consistent for most items; only minor modifications were suggested. (See *Measurement Error Studies at the National Center for Education Statistics*, NCES 97–464.)

Data Comparability

The NHES data can be compared with estimates from several other large-scale data collections, as described below.

Comparisons of methodology with other household surveys. For analysts wanting to compare the NHES surveys with another household survey, the Survey of Income and Program Participation (SIPP)—a longitudinal household survey conducted by the U.S. Bureau of the Census—provides an appropriate comparison. The first wave of data collection in SIPP is always done by personal visit to the household. Subsequent data collection is conducted primarily by telephone but may also be done in person. The response rates for SIPP are much higher than those that could be expected using a random-digit-dialing screening sample, as in the NHES program. With personal interviews, there are more opportunities to obtain participation (including activities such as speaking with neighbors), and it is easier to demonstrate the importance of the sampled person's cooperation. It should be noted that, while the difference in response rates is largely the result of the different modes of sampling and data collection, the Census Bureau's response rates are generally higher than those achieved by other collection organizations.

Comparisons of topical data. Specific data from NHES surveys can be compared with data from several other surveys, as described below.

Early childhood education. Over the years, several NHES surveys have collected similar information in early childhood education: ECPP-NHES:2001, ECPP-NHES:1995, ECE-NHES:1991, and SR-NHES:1993. These data can be compared with data from three other surveys. The *Current Population Survey (CPS)—October Education*

Supplement (conducted by the U.S. Bureau of the Census) collects information on nursery school enrollment. (See chapter 26.) CPS estimates of participation in early childhood programs and estimates of retention in early grades can be compared with NHES estimates. In addition, the 1990 CPS—October Education Supplement replicated several NHES items on home activities in which parents engage with their children. NHES data can also be compared with the *National Health Interview Survey Child Health Supplement of 1988* (conducted by the National Center for Health Statistics), which collected information on participation in child care and early childhood education programs and extensive information on the health status of children. Finally, *SIPP* (described above) periodically includes a supplement that collects information on the child care and early childhood program participation of children of mothers who are employed or enrolled in school or job training.

Before- and after-school programs and activities. ASPA-NHES:2001 covered some topics addressed in previous years by other NHES surveys. Parent-NHES:1999 and PFI/CI-NHES:1996 both collected information on school contacts with households about children. Parent-NHES:1999 also collected information on type of care and basic statistics on after-school program participation. Basic enrollment totals and demographic characteristics, as well as public and private school enrollment, can be compared with *CPS* estimates.

Adult education. Both NHES surveys (AELL-NHES:2001, AE-NHES:1999, AE-NHES:1995, and AE-NHES:1991) and *CPS* provide estimates of adult education participation. (See chapter 26.) CPS collected information on adult education participation every 3 years from 1969 through 1984. The 1992 CPS also included a brief set of questions on adult education that replicated items used to estimate the Adult Education participation rate in AE-NHES:1991.

School safety and discipline. Estimates from SS&D-NHES:1993 can be compared with three other surveys. *Monitoring the Future* (conducted by the National Institute on Drug Abuse) gathers information annually on the prevalence and incidence of the illicit drug use of 12th graders. In addition, it contains questions designed to describe and explain changes in many important values, behaviors, and lifestyle orientations of American youth. The *School Crime Supplements* of the 1989 and 1995 *National Crime Victimization Surveys* (conducted by the U.S. Department of Justice, Bureau of Justice Statistics) provide detailed information on personal crimes of violence and theft that were committed inside a school building or on school property. Finally, the NCES *National Education Longitudinal Study of 1988 (NELS:88)* provides data on educational issues such as school environment issues, school discipline issues, victimization at school, and drug and alcohol education. (See chapter 6.)

Parent involvement in education. Estimates from PFI/CI-NHES:1996 can be compared with data from *NELS:88*. (See chapter 6.) Data analysts may wish to examine NELS:88 data in conjunction with the PFI estimates on school contacts to parents (by parent report) and frequency of parents helping the child with his or her homework.

Civic involvement and other characteristics. Estimates from the NHES Adult and Youth Civic Involvement surveys can be compared with seven other surveys. The *1995 CPS—October Education Supplement* included sets of items measuring the percentage distribution of the adult population, age and sex of the adult population, household income distributions, and race/ethnicity by highest level of education. (See chapter 26.) The *1992 National Adult Literacy Survey* collected data on adults' activities in daily life that require English literacy skills. (See chapter 23.) Areas common to the *1994 General Social Survey* and ACI-NHES:1996 include organizational membership, various political or civic activities, and attitudes about freedom of speech. The *National Election Study* collects data on voting, public opinion, and political participation and knowledge during election years. Several items addressing political knowledge in ACI-NHES:1996 were drawn from the National Election Study and can be used for direct comparisons. The *Citizens' Political and Social Participation Survey* measures the extent and variety of voluntary social and political activity among Americans and the causes of that engagement. *The Washington Post/Kaiser Family Foundation/Harvard University Survey Project* provides information on public knowledge, perceptions, and attitudes about the role of American government. Finally, the *National Survey of High School Seniors* elicits detailed information on political and relevant nonpolitical matters so that parent-child similarities and differences can be assessed.

6. CONTACT INFORMATION

For content information on NHES, contact:

Chris Chapman
Phone: (202) 502–7414
E-mail: chris.chapman@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

7. METHODOLOGY AND EVALUATION REPORTS

General

National Household Education Surveys of 2001:Data File User's Manual (forthcoming), by M. Collins, J. Montaquila, M.J. Nolin, K. Kim, B. Kleiner, T. Waits, C. Chapman, and K. Chandler. Washington, DC.

National Household Education Survey of 1999: Methodology Report, NCES 2000–078, by M.J. Nolin, J. Montaquila, P. Nicchitta, K. Kim, B. Kleiner, J. Lennon, C. Chapman, S. Creighton, and S. Bielick. Washington, DC: 2000.

The 1995 National Household Education Survey: Adult Education Course Coding Manual, NCES Working Paper 97–19, by K. Kim and M.A. Collins. Washington, DC: 1997.

The 1995 National Household Education Survey: Adult Education Course Code Merge Files User's Guide, NCES Working Paper 97–20, by K. Kim and J. Brown. Washington, DC: 1997.

An Overview of the National Household Education Survey: 1991, 1993, 1995, and 1996, NCES 97–448, by M.J. Nolin, M. Collins, and J.M. Brick. Washington, DC: 1997.

Survey Design

Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey (NHES:95), NCES Working Paper 97–08, by Mary A. Collins, J.M. Brick, L.S. Loomis, P.G. Nicchitta, and S. Fleischman. Washington, DC: 1997.

Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey, NCES Working Paper 97–35, by N. Vaden-Kiernan, P.G. Nicchitta, J. Montaquila, and M.A. Collins. Washington, DC: 1997.

Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93), NCES Working Paper 97–04, by J.M. Brick, M.A. Collins, M.J. Nolin, E. Davies, and M.L. Feibus. Washington, DC: 1997.

Effectiveness of Oversampling Blacks and Hispanics in the NHES Field Test, NCES 92–104, by L. Mohadjer. Washington, DC: 1992.

An Experiment in Random-Digit-Dial Screening, NCES 98–255, by J.M. Brick and M. Collins. Washington, DC: 1997.

Feasibility of Conducting Followup Surveys in the National Household Education Survey, NCES 97–335, by J.M. Brick, M. Collins, and E. Davies. Washington, DC: 1997.

Multiplicity Sampling for Dropouts in the NHES Field Test, NCES 92–102, by J.M. Brick. Washington, DC: 1992.

Overview of the NHES Field Test, NCES 92–099, by J.M. Brick, C.A. Celebuski, and M.A. Collins. Washington, DC: 1992.

An Overview of Response Rates in the National Household Education Survey: 1991, 1993, 1995, and 1996, NCES 97–948, by J.M. Brick and M. Collins. Washington, DC: 1997.

Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93), NCES Working Paper 97–05, by J.M. Brick, E. Tubbs, M.A. Collins, and M.J. Nolin. Washington, DC: 1997.

Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95), NCES Working Paper 97–06, by J.M. Brick and P. Broene. Washington, DC: 1997.

Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey, NCES Working Paper 97–40, by J. Montaquila and J.M. Brick. Washington, DC: 1997.

Data Quality and Comparability

Comparison of Estimates from the 1995 National Household Education Survey (NHES:95), NCES Working Paper 96–30, by K. Kim, L.S. Loomis, and M.A. Collins. Washington, DC: 1996.

Comparison of Estimates from the 1993 National Household Education Survey (NHES:93), NCES Working Paper 97–34, by M.A. Collins, L. Branden, M.J. Nolin, E. Davies, and L. Loomis. Washington, DC: 1997.

Estimation of Response Bias in the NHES:95 Adult Education Survey, NCES Working Paper 96–13, by M. Brick, K. Kim, M.J. Nolin, and M.A. Collins. Washington, DC: 1996.

Measurement Error Studies at the National Center for Education Statistics, NCES 97–464, by S. Salvucci, E. Walter, V. Conley, S. Fink, and M. Saba. Washington, DC: 1997.

Measuring Participation in Adult Education, NCES 97–341, by M.A. Collins, J.M. Brick, and K. Kim. Washington, DC: 1997.

National Household Education Survey of 1993: Adjusting for Coverage Bias Using Telephone Service Interruption Data, NCES 97–336, by J.M. Brick, S. Keeter, J. Waksberg, and B. Bell. Washington, DC: 1997.

The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component, NCES Working Paper 96–14, by J.M. Brick, J. Wernimont, and M. Montes. Washington, DC: 1996.

Proxy Reporting of Dropout Status in the NHES Field Test, NCES 92–103, by L. Mohadjer and J.M. Brick. Washington, DC: 1992.

Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey, NCES Working Paper 97–38, by J.M. Montaquilla, J.M. Brick, and S.P. Brock. Washington, DC: 1997.

Reinterview Results for the School Safety and Discipline and School Readiness Components, NCES Working Paper 97–339, by J.M. Brick, L. Rizzo, and J. Wernimont. Washington, DC: 1997.

Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93), NCES Working Paper 97–02, by J.M. Brick, E. Tubbs, M.A. Collins, M.J. Nolin, D. Cantor, K. Levin, and Y. Carnes. Washington, DC: 1997.

Telephone Undercoverage Bias of 14- to 21- Year-Olds and 3- to 5-Year-Olds, NCES 92–201, by J.M. Brick and J. Burke. Washington, DC: 1992.

Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2- Year-Olds in the 1995 National Household Education Survey (NHES:95), NCES Working Paper 96–29, by J.M. Brick. Washington, DC: 1996.

Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey, NCES Working Paper 97–39, by J.M. Montaquilla, J.M. Brick, and S.P. Brock. Washington, DC: 1997.

Use of Cognitive Laboratories and Recorded Interviews in the National Household Education Survey (NHES), NCES 96–332, by M.J. Nolin. Washington, DC: 1997.

Chapter 26: Current Population Survey (CPS)—October and September Supplements

1. OVERVIEW

The Current Population Survey (CPS) is a monthly survey of about 50,000 house holds conducted by the Bureau of the Census, part of the U.S. Department of Commerce, for the Bureau of Labor Statistics (BLS), U.S. Department of Labor. The "Basic CPS" collects data about the employment, unemployment, and other characteristics of the civilian noninstitutional population in the United States; it excludes military personnel and their families living on post, inmates of institutions, and homes for the aged. Since the mid-1960s, NCES has sponsored the October Supplement to the CPS to capture information on school enrollment status and related topics for household members 3 years old and older, thus providing current estimates of school enrollment, as well as of the social and economic characteristics of students. Beginning in September 2001 NCES, in conjunction with several other federal agencies, began cosponsoring an annual survey about household and individual use of computers and the Internet. Prior to this point, computer and Internet items had been occasionally added to various CPS monthly supplements, including the October supplements.

TWO ANNUAL SUPPLEMENTS TO THE CPS

CPS Supplements collects data on household members 3 years old and over:

- School enrollment status
- Availability and use of computers and the Internet at school, home, and work

Purpose

The October Supplement is designed to collect information on the school enrollment of household members in any type of public, parochial, or other private school in the regular school system. Such schools include nursery schools, kindergartens, elementary schools, high schools, colleges, universities, and professional schools. The September Supplement is designed to collect information on the availability and use of computers and the Internet at school, home, and work.

Components

The October and September Supplements are components of CPS. The information collected is described below. An adult member of each household provides information for all members of the household.

October Supplement. The October Supplement collects information on school enrollment status and educational attainment of household members 3 years old and over, including highest grade completed, level and grade of current enrollment, attendance status, number and type of courses, degree or certificate objective, and type of organization offering instruction for each member of the household. A dozen core questions on the interview instrument for the October Supplement have remained unchanged since 1967. Since 1987, additional questions have been included on business, vocational, technical, secretarial, trade, and correspondence courses; on the grade the student was attending last year; on the calendar year that the student received his/her most recent degree; on whether or not the student completed high school by means of an equivalency test (such as the GED); and on whether or not children aged 3 to 5 are enrolled in any kind of nursery school, kindergarten, or elementary school. From time to time, additional items address such topics as private school tuition, adult education, vocational education, computer usage, and student mobility.

September Supplement. The September Supplement collects information on computer and Internet use, including whether there is a personal computer, laptop, or WebTV in the household; the number of computers or laptops; whether the newest is owned or leased, and by whom; when the newest computer was obtained; whether computers are used by students in school; and if computers are used by students for school assignments. The questions on Internet include use of Internet from the home; whether household members connect to the Internet via a computer or WebTV; the main reason for stopping Internet service if they have done so; how the Internet connection is paid for and how much is paid; which Internet service provider is used; whether long distance charges are paid to connect to the Internet service provider; how household members use the Internet including whether the Internet is used for school assignments; whether household members use the Internet outside the home including where they use it; and how concerned household members are that personal information provided to an Internet service provider may not be kept confidential.

Basic CPS. The Basic CPS collects monthly data on household membership, household characteristics, demographic characteristics, and labor force participation of the civilian noninstitutional population 16 years of age and over. The Basic CPS is collected each month from a probability sample of approximately 50,000 occupied households.

Periodicity

The October and September Supplements to the CPS are annual supplements. The Basic CPS is conducted monthly.

2. USES OF DATA

The October Supplement provides important education data to policymakers and researchers on school enrollment and educational attainment. Data from the October Supplement, together with data from the Basic CPS and the March Supplement, provide the basis for descriptive and analytic reports that portray the social and economic characteristics of students in relation to the specifics of their school enrollment. From these sources it is possible to derive retention, completion, and graduation rates, as well as high school dropout rates. Some of the October Supplements also provide policy-relevant data on private school tuition, adult education, vocational education, early childhood education, and student mobility.

The data provided by the September Supplement allows policymakers and researchers to analyze computer access and Internet use by various demographic and geographic segments of the population. Policymakers will use statistics from this supplement to come up with programs and policies that would make computer technology and the Internet as accessible as possible for as many Americans as possible.

3. KEY CONCEPTS

Some of the key concepts in the CPS October Supplement are defined below. For additional terms relevant to the October Supplement, as well as to the Basic CPS, refer to *School Enrollment—Social and Economic Characteristics of Students* (U.S. Department of Commerce, Bureau of the Census, Current Population Reports P20–413, by Robert Kominski. Washington, DC: 1987). The definition of the Internet given to respondents is also provided below.

Household. All persons who occupy a housing unit. A house, an apartment or other group of rooms, or a single room, is regarded as a housing unit when it is occupied or intended for occupancy as separate living quarters, that is, when the occupants do not live and eat with any other persons in the structure and there is direct access from the outside or through a common hall.

School Enrollment. Anyone who has been enrolled at any time during the current term or school year in any type of public, parochial, or other private school in the regular school system. Such schools include nursery schools, kindergartens, elementary schools, high schools, colleges, universities, and professional schools. Attendance may be either full-time or part-time, during the day or night. Regular schooling is that which may advance a person toward an elementary or high school diploma, or a college, university, or professional school degree. Enrollment is excluded if in schools that are not in the regular school system or that do not advance students to regular school degrees (e.g., enrollment in trade schools, business colleges, and schools for the mentally handicapped).

Level of School. Nursery school, kindergarten, elementary school (first to eighth grades), high school (9th to 12th grades), and college or professional school. The last group includes graduate students in colleges or universities. Persons enrolled in elementary, middle school, intermediate school, or junior high school through the eighth grade are classified as in elementary school. All persons enrolled in 9th through 12th grade are classified as in high school.

Nursery School. A group or class that is organized to provide educational experiences for children during the year or years preceding kindergarten. This includes Head Start programs or similar programs sponsored by local agencies to provide preschool education to young children.

Public or Private School. A public school is defined as any educational institution operated by publicly elected or appointed school officials and supported by public funds. Private schools include educational institutions established and operated by religious bodies, as well as those that are under other private control. In cases where enrollment is in a school or college that is both publicly and privately controlled or supported, enrollment is counted according to whether it is primarily public or private.

Modal Grade. For descriptive and analytic purposes, enrolled persons are classified according to their relative progress in school; that is, whether the grade or year in which they were enrolled was below, at, or above the modal (or typical) grade for persons of their age at the time of the survey. The modal grade is the year of school in which the largest proportion of students of a given age is enrolled.

Vocational School Enrollment. Vocational school enrollment includes enrollment in business, vocational, technical, secretarial, trade, and correspondence courses not counted as regular school enrollment and not for recreation or adult education classes.

Educational Attainment. Highest level of school a person has completed or highest degree a person has received.

Internet. The Internet is an electronic network that connects more than 300 million users across the world. These users are linked to the Internet by computer or various telecommunication devices, and use it to communicate through e-mail, to obtain information, to purchase products, etc.

4. SURVEY DESIGN

Target Population

All household members aged 3 years and older in the civilian noninstitutional population of the 50 states and the District of Columbia. Excludes military personnel and their families living on post, inmates of institutions, and homes for the aged.

Sample Design

The Basic CPS is based upon a probability sample of about 50,000 housing units. Each month, interviewers contact the sampled units to obtain basic demographic information on all persons residing at the address and detailed labor force information on all persons aged 15 or over. To improve the reliability of estimates of month-to-month and year-to-year change, eight panels are used to rotate the sample each month. A sample unit is interviewed for 4 consecutive months, and then, after an 8-month rest period, for the same 4 months a year later. Every month, a new panel of addresses, or one-eighth of the total sample, is introduced. Thus, in a particular month, one panel is being interviewed for the first time, one panel for the second, ..., and one panel for the eighth and final time.

The first stage sample selection is carried out in three major steps: definition of the PSUs; stratification of the PSUs within each state; and selection of the sample PSUs in each state. The CPS national design as of January 1996 contains 754 stratification PSUs. Using a Maximum Overlap procedure, one PSU is selected per stratum with probability proportional to its 1990 population. This procedure uses mathematical programming techniques to maximize the probability of selecting PSUs that are already in sample while maintaining the correct overall probabilities of selection.

The second stage of the CPS sample design is the selection of sample housing units within PSUs. These ultimate sampling unit (USU) clusters consist of a geographically compact cluster of approximately four addresses, corresponding to four housing units at the time of the census.

Each month, about 59,000 housing units are assigned for data collection, of which about 50,000 are occupied and thus eligible for interview. The remainder are units found to be destroyed, vacant, converted to nonresidential use, containing persons whose usual place of residence is elsewhere, or ineligible for other reasons. Of the 50,000 housing units, about 6.5 percent are not interviewed in a given month due to temporary absence (vacation, etc.), other failures to make contact after repeated attempts, inability of persons contacted to respond, unavailability for other reasons, and refusals to cooperate (about half of the noninterviews). In 1999, information was obtained each month on about 94,000 persons 16 years of age or older and on approximately 29,000 persons under the age of 16.

Data Collection and Processing

The U.S. Bureau of the Census is the collection agent for the CPS and its supplements. Additional details on data collection and processing are provided in *The Current Population Survey: Design and Methodology* (Technical Paper 63).

Reference dates. The reference period for the October Supplement is the current school year, which is assumed to be in progress in the interview month of October. The reference period for the labor force questions on the underlying Basic CPS is the week that contains the 12th of the month. The reference period for the September Supplement is the current year.

Data collection. Each month, Bureau of the Census field representatives attempt to collect data from the sample units during the week containing the 19th of the month. For the first month-in-sample interview, the interviewer visits the sample address to determine if the sample unit exists, if it is occupied, and if some responsible adult will provide the necessary information. If someone at the sample unit agrees to the interview, the interviewer uses a laptop computer to administer the interview. In most cases, the interviewer conducts subsequent interviews by telephone (use of telephone interviewing must be approved by the respondent) and does not actually visit the sample unit again until the fifth month-in-sample interview, the first interview after the 8-month resting period. Fifth-month households are more likely than any other household to be a replacement household; that is, a household in which all the previous month's residents have moved out and been replaced by an entirely different group of residents. However, any person can change his/her household status during the time in sample: a person who leaves the household is deleted from the roster; a person who moves into the household is added to the roster.

Most month-in-sample 2 through 4 and 6 through 8 interviews are conducted by telephone (e.g., 87 percent in December 1996). Interviewers continue to visit households without telephones, with poor English-language skills, or which decline a telephone interview.

The interview begins with questions about the housing unit and the people who consider this address their usual residence. Basic demographic information is collected for each household member. Labor force information is collected for each civilian 15 years of age or older, although the data for 15-year-olds are not used in official BLS estimates. After the labor force information has been collected for all eligible household members, supplemental questions particular to that month's interview may be asked of specific family members or the entire household.

Editing. Completed interviews are electronically transmitted to a central processor where the responses are edited for consistency and various codes are added. The edits effectively blank out all entries in inappropriate questions and ensure that all appropriate questions have valid entries.

Estimation Procedures

Weighting is used in the CPS to adjust for sampling and unit nonresponse, and imputation is used to adjust for item nonresponse.

Weighting. For the Basic CPS, the estimation procedure involves weighting the data from each sample person by the inverse of the probability of the person's housing unit being in the sample. With some exceptions, sample persons within the same state have the same probability of selection. The CPS uses raking ratio estimation to derive the weights used to tabulate total U.S. and state estimates. The goal is to control the survey estimates of the population in specific subgroups to independently derived estimates of the civilian noninstitutional population in the 50 states and the District of Columbia. In addition, household and family weights provide a basis for household-level estimates and estimates for married couples living in the same household.

For all CPS data files, a final weight is prepared and used to compute the monthly labor force status estimates. The final weight, which is the product of several adjustments, including a nonresponse adjustment, is used to produce estimates for the various characteristics covered in the full monthly CPS. This weight is constructed from the basic weight for each person, which represents the probability of selection for the survey. For supplements, such as the October Supplement, separate data processing is required, not only to edit responses for consistency and impute for missing values, but also to incorporate special weighting procedures to account for the fact that the supplement is targeting a special universe, such as school-age children, in contrast to the working-age labor force emphasis of the Basic CPS. However, there is no supplement weight associated with the October 1998 School Enrollment Supplement.

Starting with the data collected in the October 1994 CPS, independent estimates are based on civilian noninstitutional population controls for age, race, and sex established by the 1990 decennial census and adjusted for an undercount of about 1.6 percent. These independent estimates are based on statistics from decennial censuses; statistics on births, deaths, immigration, and emigration; and statistics on the size of the Armed Forces.

Imputation. When a response is not obtained for a particular data item, or an inconsistency in reported items is detected, an imputed response is entered in the field. Note that edits are run in a deliberate sequence: demographic variables are edited first because several of those variables are used to allocate missing values in the other modules, and the labor force module is edited next since labor force status and related items are used to impute missing values for industry and occupation codes and so forth.

CPS edits use three imputation methods: relational imputation, longitudinal edits, and hot-deck imputation. Relational imputation infers the missing value from other characteristics on the person's record or within the household. Longitudinal edits are used primarily in the labor force edits. If a question is blank and the record is in the overlap sample, the edit looks at the previous month's data to determine whether the person had responded then for that item. If so, the previous month's entry is assigned; otherwise, the item is assigned a value using the appropriate hot deck. The hot-deck method assigns a value from a record with similar characteristics. Hot decks are always defined by age, race, and sex. Other characteristics used in hot decks vary depending on the nature of the question being referenced. The imputation procedure is performed one item at a time. In a typical month, the imputation rate for demographic items is less than 1 percent. The rates for labor force items are slightly over 1 percent. Over all earnings items, the imputation rate is near 10 percent, with some items having much higher and others much lower nonresponse rates. In October 1998, the imputation rate for the basic school enrollment items ranged from 4–7 percent per item.

Future Plans

The October Supplement will always include the traditional enrollment questions; questions on other topics will be added as occasion warrants. For example, the October Supplement for 1997 included questions on computer use, and the October Supplement for 1999 included questions on English language proficiency, disabilities, and grade retention. The 2000 and 2001 October Supplements included only the enrollment questions. Plans for additional questions in future years have yet to be determined. The September Supplement will continue to include questions about computer and Internet access and use for the foreseeable future with some topical flexibility to account for the rapidly changing computer and telecommunications environment.

5. DATA QUALITY AND COMPARABILITY

Sampling Error

Although the estimation methods used in the CPS do not produce unbiased estimates, biases for most estimates are believed to be small enough so that these confidence interval statements are approximately true. Standard error estimates computed using generalized variance functions are provided in *Employment and Earnings* and other BLS publications. Using replicate variance techniques, standard error estimates are generated. As computed, these standard error estimates reflect contributions not only from sampling error but also from some types of nonsampling error, particularly response variability. Because replicate variance techniques are somewhat cumbersome, simplified formulas called generalized variance functions (GVFs) have been developed for various types of labor force characteristics. The GVF can be used to approximate an estimate's standard error, but this only indicates the general magnitude of its standard error rather than a precise value.

Nonsampling Error

Although the full extent of nonsampling error in the CPS is unknown, special studies have been conducted to quantify some of the possible sources. The effect of nonsampling error should be small on estimates of relative change, such as month-to-month change. Estimates of monthly levels would be more severely affected by nonsampling error.

Coverage error. Undercoverage in the CPS results from missed housing units and missed persons within sample households. The CPS covers about 92 percent of the decennial census population (adjusted for the undercount). It is known that the CPS undercoverage varies with age, sex, race, and Hispanic origin. Generally, undercoverage is larger for men than for women and larger for Blacks, Hispanics, and other races than for Whites. Ratio adjustment to independent age/sex/race/origin population controls, as described previously, partially corrects for the biases due to survey undercoverage. However, biases exist in the estimates to the extent that missed persons in missed households or missed persons in interviewed households have different characteristics than interviewed persons in the same age/sex/race/origin group.

The independent population estimates used in the estimation procedure may be a source of error although, on balance, their use substantially improves the statistical reliability of many of the figures. Errors may arise in the independent population estimates because of underenumeration of certain population groups or errors in age reporting in the 1990 census (which serves as the base for the estimates) or similar problems in the components of population change (mortality, immigration, etc.) since that date.

Nonresponse error.
Unit nonresponse. Unit nonresponse may have a number of components. A respondent may refuse to participate in the survey, may not be capable of completing the interview, or may not be available to the interviewer during the specified survey period. If the entire household does not participate, this situation is referred to as a "Type A noninterview." There is also another type of (partial) unit nonresponse, namely that one or more individual persons within the household refuses to be interviewed. This is not a major problem in the CPS since any responsible adult may be able to report information for other persons as a proxy reporter. There are other variations on unit nonresponse; detailed consideration of these may be found in The Current Population Survey: Design and Methodology (Technical Papers 40 and 63). For the October 2000 basic CPS, the nonresponse rate was 6.8 percent and for the school enrollment supplement the nonresponse rate was an additional 3.1 percent for a total supplement nonresponse rate of 9.7 percent.

Item nonresponse. Although an imputation procedure is implemented for item nonresponse in the CPS, there is no way of assuring that the errors of item imputation will balance out and that any potential bias has been avoided.

Measurement error. The main sources of nonsampling variability in the responses to the October Supplement are those inherent in the survey instrument. The question of current enrollment may not be answered accurately for various reasons. Some respondents may not know current grade information for every student in the household, a problem especially prevalent for households with members in college or in nursery school. Confusion over college credits or hours taken by a student may make it difficult to determine the year in which the student is enrolled. Problems may occur with the definition of nursery school (a group or class organized to provide educational experiences for children), where respondents' interpretations of "educational experiences" vary.

6. CONTACT INFORMATION

For content information about the September and October Supplements, contact:

NCES Contact:
Chris Chapman
Phone: (202) 502–7414
E-mail: chris.chapman@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

Census Bureau Contact:
Jennifer Day
Phone: (301) 763–2464
E-mail: jday@census.gov

Mailing Address:
Education and Social Stratification Branch
Population Division
Bureau of the Census
U.S. Department of Commerce
Washington, DC 20233

7. METHODOLOGY AND EVALUATION REPORTS

General

U.S. Department of Commerce, Bureau of the Census. *The Current Population Survey: A Report on Methodology*, Technical Paper 7, by J. Steinberg, T.B. Jabine, and L. Pritzker. Washington, DC: 1963.

U.S. Department of Commerce, Bureau of the Census, and U.S. Department of Labor, Bureau of Labor Statistics. *Concepts and Methods Used in Labor Force Statistics Derived From the Current Population Survey*, Current Population Reports, Special Studies Series P23, No. 62. Washington, DC: 1976.

U.S. Department of Labor, Bureau of Labor Statistics. *BLS Handbook of Methods*. Washington, DC: 1997.

Survey Design

Frankel, L.R. and Stock, J.S. "On the Sample Survey of Unemployment." *Journal of the American Statistical Association*, 37 (March 1942): 77–80.

Hansen, M.H. and Hurwitz, W.N. "On the Theory of Sampling From Finite Populations." *Annals of Mathematical Statistics*, XIV (1943): 333–362.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. *Sample Survey Methods and Theory Vol. I, Methods and Applications*. New York: John Wiley and Sons, 1953.

Hansen, M.H., Hurwitz, W.N., Nisselson, H., and Steinberg, J. "The Redesign of the Census Current Population Survey." *Journal of the American Statistical Association*, 50 (September 1955): 701–719.

U.S. Department of Commerce, Bureau of the Census. *The Current Population Survey: Design and Methodology*, Technical Paper 40. Washington, DC: 1978.

U.S. Department of Commerce, Bureau of the Census. *The Current Population Survey: Design and Methodology*, Technical Paper 63. Washington, DC: 2000.

U.S. Department of Commerce, Bureau of the Census. *The X11 Variant of the Census Method II Seasonal Adjustment Program*, Technical Paper 15, by J. Shiskin, A.H. Young, and J.C. Musgrave. Washington, DC: 1965.

Data Quality and Comparability

U.S. Department of Commerce, Office of Federal Statistical Policy and Standards. *An Error Profile: Employment as Measured by the Current Population Survey*, Statistical Policy Working Paper 3. Washington, DC: 1978.

Chapter 27: Fast Response Surveys

NCES has established two survey systems to collect time-sensitive, issue-oriented data quickly and with minimum response burden. The Fast Response Survey System (FRSS) focuses on collecting data at the elementary and secondary school level. The Postsecondary Education Quick Information System (PEQIS) collects data at the postsecondary level. These systems are used to meet the data needs of Department of Education analysts, planners, and decision makers when information cannot be obtained quickly through traditional NCES surveys.

TWO FAST RESPONSE SYSTEMS:

- Fast Response Survey System (FRSS)—80 surveys since 1975
- Postsecondary Education Quick Information System (PEQIS)—12 surveys since 1991

1. FAST RESPONSE SURVEY SYSTEM (FRSS)

Overview

The Fast Response Survey System (FRSS) was established in 1975 to collect small amounts of data on key education issues within a relatively short time frame. From 1975 to 1990, FRSS collected data at all educational levels. Since the Postsecondary Education Quick Information System (PEQIS) was established in 1991, FRSS surveys have been limited to elementary and secondary school issues. To date, nearly 80 surveys have been conducted under FRSS. Topics have ranged from racial and ethnic classifications at state and school levels to the availability and use of resources such as advanced telecommunications and libraries. Additionally, data have been collected on education reform, violence and discipline problems, parental involvement, curriculum placement and arts education, nutrition education, teacher training and professional development, vocational education, children's readiness for school, and the perspectives of school district superintendents, principals, and teachers on safe, disciplined, and drug-free schools.

Data from FRSS surveys are representative at the national level, drawing from a universe that is appropriate for each study. Since 1992, FRSS has generally collected data from public and private elementary and secondary schools, elementary and secondary school teachers and principals, and public and school libraries. In its earlier years, FRSS also collected data from state education agencies and other educational organizations and participants, including local education agencies.

Sample Design

The sampling frame for FRSS surveys is typically the Common Core of Data (CCD) public school (or agency) universe. (See chapter 2.) The following variables are usually used for stratification or sorting within primary strata: instructional level (elementary school, middle school and high—secondary/combined—school); size of enrollment; locale (city, urban fringe, town, rural); geographical region (Northeast, Southeast, Central, West); percent minority enrollment; and/or poverty status (based on eligibility for free or reduced-price lunch). The allocation of the samples to the primary strata is intended to ensure that the sample sizes are large enough to permit analyses of the questionnaire for major subgroups. Within primary strata, the sample sizes are frequently allocated to the substrata in rough proportion to the aggregate square root of the size of enrollment of schools in the substratum. The use of the square root of enrollment to determine the sample allocation is considered reasonably efficient for estimating both school-level characteristics and quantitative measures correlated with enrollment.

FRSS survey samples are sometimes constructed from the Private School Universe Survey (PSS). (See chapter 3.) The sample usually consists of *regular* private elementary, middle, secondary, and combined schools, with a private school being defined as a school not in the public system that provides instruction for any of grades 1–12 (or comparable ungraded levels) where the instruction was not provided in a private home. The following variables may be used for stratification or sorting within primary strata: instructional level (elementary, secondary, combined), affiliation (Catholic, other religious, and nonsectarian), school size, geographic region, locale, and percent minority enrollment. Schools are generally selected from each primary stratum with probabilities proportional to the weight reflecting the school's probability of inclusion in the area sample.

Other sources may serve as sampling frames, depending on the needs of the survey. For example, for *Participation of Migrant Students in Title I Migrant Education Program (MEP) Summer-Term Projects*, the districts and other entities serving migrant students were selected from the U.S. Department of Education's 1995–96 Migrant Education Program Universe file.

Some FRSS surveys use a two-stage sampling process. For example, the *Teacher Survey on Safe, Disciplined, and Drug-Free Schools* and the *Public School District Survey on Safe, Disciplined, and Drug-Free Schools* were administered concurrently with the *Principal Survey on Safe, Disciplined, and Drug-Free Schools*. Both the Teacher and Public School District surveys had a two-stage sampling process. The schools were selected during the first stage. The second stage of sampling for the Teacher Survey involved obtaining lists of teachers from the selected schools. The second stage of sampling for the Public School District Survey identified the districts to be included in the survey. Districts consisting of two or more schools had multiple chances of selection. The overall probability of selecting a district was equal to the probability that any of its constituent schools was selected for the principal's survey.

Before PEQIS was established, FRSS was sometimes used to examine postsecondary issues. For example, the 1990 *Survey of Remedial/Developmental Studies in Institutions of Higher Education* targeted institutions of higher education (IHEs) that served freshmen and were accredited at the college level by an association or agency recognized by the U.S. Secretary of Education. The sampling frame was the universe file of the Higher Education General Information System (HEGIS) Fall Enrollment and Compliance Report of Institutions of Higher Education of 1983–84. (Note that HEGIS has since been replaced by the Integrated Postsecondary Education Data System —IPEDS—see chapter 14.) The universe of colleges and universities was stratified by type of control, type of institution, and enrollment size. Within strata, schools were selected at uniform rates, but the sampling rates varied considerably from stratum to stratum.

Data Collection and Processing

Most FRSS surveys are self-administered questionnaires that are mailed to the respondents with telephone and fax follow up. A few have been telephone surveys, including one which used Random Digit Dialing (RDD) techniques. FRSS questionnaires are pretested and efforts are made to check for consistency of interpretation of questions and to eliminate ambiguous items before fielding the survey.

Data are keyed with 100 percent verification. To check the data for accuracy and consistency, questionnaire responses undergo both manual and machine editing. Cases with missing or inconsistent items are recontacted by telephone.

Westat has served as the contractor for all surveys.

Weighting

The response data are weighted to produce national estimates. The weights are designed to adjust for the variable probabilities of selection and differential nonresponse. Out-of-scope units are deleted from the initial sample before weighting and analysis. In the case of two-stage sampling—for example, in the *Teacher Survey on Safe, Disciplined, and Drug-Free Schools*—the weights used to produce national estimates are equal to the reciprocal of the product of the probability of selecting the school and the probability of selecting the teacher, multiplied by an adjustment to account for school and teacher nonresponse.

Imputation

Because item nonresponse rates in FRSS surveys are low, imputation has only been performed for one survey—the 1990 Survey of Remedial/Developmental Studies in Institutions of Higher Education. In that instance, seven items required imputation: percent enrolled in remedial reading, writing, mathematics courses (three items); percent passing remedial reading, writing, mathematics courses (three items); and percent enrolled in remedial courses in reading, writing, or mathematics (one item). For the first six items, a sequential hot-deck imputation procedure was used. Imputations for the seventh item—total percentage of freshmen enrolled in one or more remedial courses in reading, writing, or mathematics—were restricted by the maximum and minimum values for the percentage enrolled in each of the individual subjects (remedial reading, writing, and mathematics). Because of these restrictions, it was decided to impute the midpoint (i.e., median) between the minimum and maximum values. The imputed values for this item had a slightly larger but still statistically insignificant impact on the estimated overall average percentage of students enrolled in one or more remedial courses.

Sampling Error

FRSS estimates are based on the selected samples and, consequently, are subject to sampling variability. The standard error is a measure of the variability of estimates due to sampling. Jackknife replication is the method used to compute estimates of standard errors.

Coverage Error

FRSS surveys are subject to any coverage error present in the major NCES data files that serve as their sampling frames. Many FRSS surveys use the CCD surveys as the sampling frame. The report *Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary Education Agency Universe Survey* (NCES 97-505) found that overall coverage in the Agency Universe Survey was 96.2 percent (in a comparison to state education directories). "Regular" agencies—those traditionally responsible for providing public education—had almost total coverage in the 1994–95 survey. Most coverage discrepancies were attributed to nontraditional agencies that provide special education, vocational education, and other services. Most FRSS surveys exclude nontraditional schools. However, there is potential for undercoverage bias associated with the absence of schools built between the construction of the sampling frame and time of the FRSS survey administration. Since teacher coverage depends on teacher lists sent by the schools, teacher coverage is assumed to be good. (See chapter 2 for a description of the CCD; see relevant chapters for other NCES surveys that serve as sampling frames for FRSS surveys.)

Nonresponse Error

Unit response for most FRSS surveys is 90 percent or higher. (See the table below.) *Item nonresponse* for most

Table 12. Weighted unit response rates for several recent FRSS surveys, 1996–1999

Survey	List participation rate	Weighted first level response rate	Overall weighted response rate
National Student Service-Learning and Community Service Survey (1999)	†	93	93
Public School Teachers' Use of Computers and the Internet (1999)	*91	*91	*83
Survey on the Condition of Public School Facilities (1999)	†	91	91
Vocational Programs in Secondary Schools (1999)	†	95	95
Survey on Advanced Telecommunications in U.S. Private Schools: 1998–99	†	84	84
Participation of Migrant Students in Title I Migrant Education Program (MEP) Summer-Term Projects (1998)	†	91	91
Teacher Survey on Professional Development and Training (1997–98)	93	92	86
Principal/School Disciplinarian Survey on School Violence (1997)	†	89	89
Public School Survey on Education Reform (1996)	†	90	90
Public School Teacher Survey on Education Reform (1996)	95	90	86
Survey on Family and School Partnerships in Public Schools, K–8 (1996)	†	92	92

*Unweighted

†Not applicable

SOURCE: Alexander, Heaviside, and Farris, *Status of Education Reform in Public Elementary and Secondary Schools: Teachers' Perspectives* (NCES 1999–045). Carey, Lewis, and Farris, *Parent Involvement in Children's Education: Efforts by Public Elementary Schools* (NCES 98–032). Celebuski and Farris, *Status of Education Reform in Public Elementary and Secondary Schools: Principals' Perspectives* (NCES 98–025). Heaviside, Rowand, Williams, Farris, Burns, and McArthur, *Violence and Discipline Problems in U.S. Public Schools: 1996–97* (NCES 98–030). Lewis, Parsad, Carey, Bartfai, Farris, and Smerdon, *Teacher Quality: A Report on the Preparation and Qualifications of Public School Teachers* (NCES 1999–080). Lewis, Snow, Farris, Smerdon, Cronen, and Kaplan, *Condition of America's Public School Facilities: 1999* (NCES 2000–032). Parsad and Farris, *Occupational Programs and the Use of Skill Competencies at the Secondary and Postsecondary Levels, 1999* (NCES 2000–023). Parsad, Heaviside, Williams, and Farris, *Participation of Migrant Students in Title I Migrant Education Program (MEP) Summer-Term Projects, 1998* (NCES 2000–061). Parsad, Skinner, and Farris, *Advanced Telecommunications in U.S. Private Schools: 1998–99* (NCES 2001–037). Skinner and Chapman, *Service-Learning and Community Service in K–12 Public Schools* (NCES 1999–043). Smeardon, Cronen, Lanahan, Anderson, Iannotti, and Angeles, *Teachers' Tools for the 21st Century: A Report on Teachers' Use of Technology* (NCES 2000–102).

BEST COPY AVAILABLE

items is less than 1 percent. The weights are adjusted for unit nonresponse. As mentioned earlier, because item nonresponse rates have been low, imputation has only been implemented for one survey.

Measurement Error

Errors may result from such problems as misrecording of responses; incorrect editing, coding, and data entry; different interpretations of definitions and the meaning of questions; memory effects; the timing of the survey; and the respondent's inability to report certain data due to its recordkeeping system. One specific example of possible measurement error comes from the *Public School Survey on Education Reform* and the *Public School Teacher Survey on Education Reform*, conducted in 1996. Survey results should be interpreted carefully for the following reasons: (1) survey questions were designed to be inclusive of a wide variety of reform activities since all principals and teachers do not share the same concept of reform; (2) respondents may overreport activities in which they believe they should be engaged; and (3) the questionnaire was too brief to collect information that could assist in judging the accuracy of the respondents' reports.

Data Comparability

Some FRSS surveys are repeated so that results can be compared over time. For example, the *Survey on Advanced Telecommunications in U.S. Public Schools, K–12*, was administered annually from 1994 to 1997, and the *Survey on Advanced Telecommunications in U.S. Private Schools* was administered in 1995 and 1998–99. The 1997 *Principal/School Disciplinarian Survey on School Violence* can be compared with results from the 1991 *Principal Survey on Safe, Disciplined, and Drug-Free Schools*, although there are some sampling differences that should be taken into account. (The 1997 survey was restricted to regular elementary and secondary schools, whereas the 1991 survey also included 13 vocational education and alternative schools in the sample.) The 1990 *Survey of Remedial/Developmental Studies in Institutions of Higher Education* results updated the results from a 1983–84 FRSS survey on the same topic, and a third survey on remedial education was conducted under the PEQIS system in 1995.

Occasionally, an FRSS survey is fielded to provide data that can be compared with another NCES survey. For example, the 1996 *Survey on Family and School Partnerships in Public Schools, K–8*, was designed to provide data that could be compared with parent data in the 1996 National Household Education Survey and with the Prospects Study, a congressionally mandated study of educational growth and opportunity from 1991 to 1994.

Contact Information

For content information on FRSS, contact:

Bernard R. Greene
Phone: (202) 502–7348
E-mail: bernard.greene@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

Methodology and Evaluation Reports

Methodology discussed in technical notes to survey reports. Some recent reports are listed below.

Advanced Telecommunications in U.S. Private Schools: 1998–99, NCES 2001–037, by B. Parsad, R. Skinner, and E. Farris. Washington, DC: 2001.

College-Level Remedial Education in the Fall of 1989, NCES 91–191, by W. Mansfield, E. Farris, and M. Black. Washington, DC: 1991.

Condition of America's Public School Facilities: 1999, NCES 2000–032, by L. Lewis, K. Snow, E. Farris, B. Smerdon, S. Cronen, and J. Kaplan. Washington, DC: 2000.

Occupational Programs and the Use of Skill Competencies at the Secondary and Postsecondary Levels, 1999, NCES 2000–023, by B. Parsad and E. Farris. Washington, DC: 2000.

Parent Involvement in Children's Education: Efforts by Public Elementary Schools, NCES 98–032, by N. Carey, L. Lewis, and E. Farris. Washington, DC: 1998.

Participation of Migrant Students in Title I Migrant Education Program (MEP) Summer-Term Projects, 1998, NCES 2000–061, by B. Parsad, S. Heaviside, C. Williams, and E. Farris. Washington, DC: 2000.

Service-Learning and Community Service in K–12 Public Schools, NCES 1999–043, by R. Skinner and C. Chapman. Washington, DC: 1999.

Status of Education Reform in Public Elementary and Secondary Schools: Principals' Perspectives, NCES 98–025, by C. Celebuski and E. Farris. Washington, DC: 1998.

Status of Education Reform in Public Elementary and Secondary Schools: Teachers' Perspectives, NCES 1999–045, by D. Alexander, S. Heaviside, and E. Farris. Washington, DC: 1999.

Teacher Quality: A Report on the Preparation and Qualifications of Public School Teachers, NCES 1999–080, by L. Lewis, B. Parsad, N. Carey, N. Bartfai, E. Farris, and B. Smerdon. Washington, DC: 1999.

Teachers' Tools for the 21st Century: A Report on Teachers' Use of Technology, NCES 2000–102, by B. Smeardon, S. Cronen, L. Lanahan, J. Anderson, N. Iannotti, and J. Angeles. Washington, DC: 2000.

Violence and Discipline Problems in U.S. Public Schools: 1996–97, NCES 98–030, by S. Heaviside, C. Rowand, C. Williams, E. Farris, S. Burns, and E. McArthur. Washington, DC: 1998.

2. POSTSECONDARY EDUCATION QUICK INFORMATION SYSTEM (PEQIS)

Overview

The Postsecondary Education Quick Information System (PEQIS) was established in 1991 to quickly collect limited amounts of policy-relevant information from a nationally representative sample of postsecondary institutions. PEQIS surveys are also used to assess the feasibility of developing large-scale data collection efforts on a given topic or to supplement other NCES postsecondary surveys. To date, 12 PEQIS surveys have been completed, covering such diverse issues as distance learning, precollegiate programs for disadvantaged students, remedial education, campus crime and security, finances, services for deaf and hard of hearing students, and accommodation of disabled students.

Sample Design

PEQIS employs a standing sample (panel) of approximately 1,600 nationally representative postsecondary education institutions. Two panels have been recruited since PEQIS was established in 1991. The sampling frame for the first PEQIS panel, recruited in 1992, was the 1990–91 Integrated Postsecondary Education Data System (IPEDS) Institutional Characteristics (IC) file. (See chapter 14.) The sampling frame for the second PEQIS panel, recruited in 1996, was the 1995–96 IPEDS IC file. The PEQIS panel was reselected in 1996 to reflect changes in the postsecondary education universe since the 1992 panel was recruited. A modified Keyfitz approach was used to maximize overlap between the two panels.

Institutions eligible for the PEQIS frames for both the 1992 and 1996 panels included 2-year and 4-year (including graduate-level) postsecondary institutions, and less-than-2-year institutions of higher education. In 1992, these institutions covered the 50 states, the District of Columbia, and Puerto Rico. In 1996, institutions in Puerto Rico were excluded. There were 5,317 institutions in the 1992 sampling frame, and 5,353 institutions in the 1996 sampling frame.

The sampling frames for both PEQIS panels were stratified by instructional level (4-year, 2-year, less-than-2-year); control (public, private nonprofit, private for-profit); highest level of offering (doctor's/first professional, master's, bachelor's, less than bachelor's); total enrollment; and status as either an institution of higher education or other postsecondary institution. Within each of the strata, institutions were sorted by region (Northeast, Southeast, Central, West), whether the institution had a relatively high minority enrollment, and whether the institution had research expenditures exceeding $1 million. The 1992 sample of 1,665 institutions was allocated to the strata in proportion to the aggregate square root of *full-time-equivalent enrollment*. The 1996 sample of 1,669 institutions was allocated to the strata in proportion to the aggregate square root of *total enrollment*. For both panels, institutions within a stratum were sampled with equal probabilities of selection.

During recruitment for the 1992 panel, 50 institutions were found to be ineligible for PEQIS, primarily because they had closed or offered just correspondence courses. The final unweighted response rate at the end of PEQIS panel recruitment in spring 1992 was 98 percent (1,576 of the 1,615 eligible institutions). The weighted response rate for panel recruitment (weighted by the base weight) was 96 percent.

The modified Keyfitz approach used in 1996 resulted in 80 percent of the institutions in the 1996 panel overlapping the 1992 panel. Panel recruitment was conducted with the 338 institutions that were not part of the overlap sample. Twenty institutions were found to be ineligible for PEQIS. The final unweighted response rate for the institutions that were not part of the overlap sample was 98 percent. The final participation rate across all 1,669 institutions selected for the 1996 panel was 99.6 percent, or 1,628 out of 1,634 eligible institutions. The weighted panel participation rate (weighted by the base weight) was 99.7 percent.

Data Collection and Processing

All PEQIS surveys are mailed self-administered questionnaires. Surveys are limited to three pages of questions, with a response burden of about 30 minutes per respondent. The questionnaires are pretested and efforts are made to check for consistency of interpretation of questions and to eliminate ambiguous items before fielding the survey to all institutions in the sample.

The questionnaires are sent to institutional survey coordinators who identify the appropriate respondents for the particular survey and forward questionnaires to those persons. Nonrespondents who have not returned the survey within a set period of time are followed up by telephone. Data are keyed with 100 percent verification. To check the data for accuracy and consistency, questionnaire responses undergo both manual and machine editing. Cases with missing or inconsistent items are recontacted by telephone.

Westat has served as the contractor for all surveys.

Weighting

The response data are weighted to produce national estimates. The weights are designed to adjust for the variable probabilities of selection and differential nonresponse. Out-of-scope units are deleted from the sample before weighting and analysis.

Imputation

Item nonresponse rates in PEQIS surveys have been very low, so imputation has only been performed for two surveys. All nonresponse on the 1997–98 *Survey on Distance Education Courses Offered by Higher Education Institutions* was imputed using a combination of standard (random within class) hot-deck imputation procedures (for questions involving numbers of courses and enrollments) and/or assignment of modal values from imputation classes on the question concerning plans for distance education technologies. For the 1992 *Survey on Deaf and Hard of Hearing Students in Postsecondary Education*, the three items with the highest nonresponse rates were imputed. These items requested, respectively, the number of deaf and hard of hearing students enrolled at the institution in each of 4 academic years from 1989–90 through 1992–93; the number of such students to whom *any* special support services were provided by the institution; and the number of such students provided *specific types* of support services (sign language interpreters, oral interpreters, classroom notetakers, tutors, assistive listening devices, etc.). The imputation procedures involved a combination of standard hot-deck imputation for institutions missing data for all 4 years and, for institutions that provided data for one or more of the 4 years, application of subsequent years' data to previous years, adjusted by the average rate of change of similar institutions (based on sampling strata).

Sampling Error

Estimates are based on the selected samples and, consequently, are subject to sampling variability. The standard error is a measure of the variability of estimates due to sampling. Jackknife replication is the method used to compute estimates of standard errors.

Coverage Error

Because the frames for PEQIS surveys are constructed from IPEDS, coverage error is believed to be minimal.

Nonresponse Error

Both unit nonresponse and item nonresponse are quite low in PEQIS surveys. For the 12 surveys completed thus far, weighted *unit response* has ranged from 90 to 97 percent. *Item nonresponse* for most items in PEQIS surveys has been less than 1 percent. The weights are adjusted for unit nonresponse. As mentioned earlier, because item nonresponse rates have been low, imputation has only been implemented twice.

Measurement Error

This type of nonsampling error may result from different interpretations of survey definitions by respondents or the institution's inability to report according to survey specifications due to its recordkeeping system. Some examples of measurement error in PEQIS surveys follow.

In the 1996 *Survey on Campus Crime and Security at Postsecondary Education Institutions*, the crime statistics collected were only for occurrences of crimes committed *on* campus; the victims could be students, staff, or campus visitors. Also, these statistics only reflect crimes that were *reported* to local police agencies or to any institution official with responsibility for student and campus activities.

The 1995 *Survey on Remedial Education in Higher Education Institutions* was conducted to provide current national estimates on the extent of remediation on college

BEST COPY AVAILABLE

campuses. Institutions provided information about their remedial reading, writing, and mathematics courses offered in fall 1995. Remedial courses were defined as courses designed for college students lacking those skills necessary to perform college-level work at the level required by the institution. Thus, what constituted remedial courses varied by institution. Respondents were asked to include any courses meeting the definition, regardless of name. Some institutions refer to remedial courses as "compensatory," "developmental," or "basic skills."

In the 1994 *Survey on Precollegiate Programs for Disadvantaged Students at Higher Education Institutions*, some institutions failed to properly identify their largest precollegiate program due to the lack of a centralized information source about precollegiate programs. After data collection was completed, eight responding institutions were externally identified as having Upward Bound programs, although on the survey they reported having no precollegiate programs for the disadvantaged. It is probable that other non-Upward Bound precollegiate programs were also omitted. The failure to report having a precollegiate program may be more likely when an institution has only small, less visible programs. For similar reasons, some respondents with multiple precollegiate programs may have misidentified the largest program. However, numerous errors of this type were detected and resolved during data collection, so misidentification of the largest programs should be a relatively infrequent error. Another effect of the decentralized structure of precollegiate programs is that institutional respondents had little sense of how the largest program compared to the totality of all programs. Institutions could only compare the largest program to others of which they were aware.

The 1993 *Survey on Deaf and Hard of Hearing Students in Postsecondary Education* gathered information about the range of postsecondary institutions in which deaf and hard of hearing students enroll, the number of such students enrolled, and the support services provided to these students by the postsecondary institutions. However, institutions could only report about those students who had identified themselves to the institution as deaf or hard of hearing; thus it is likely that the survey results represent only a subset of all deaf or hard of hearing postsecondary students. Moreover, no definitions of these terms were provided to the institutions.

Data Comparability

While most PEQIS surveys are not designed specifically for comparison with other surveys, the data from some PEQIS surveys can be compared with data from other postsecondary surveys. There have been, however, two administration of the PEQIS *Survey on Distance Education Courses Offered by Higher Education Institutions.*

The 1998 *Survey on Students with Disabilities at Postsecondary Education Institutions* complements another recent NCES study on the self-reported preparation, participation, and outcomes of students with disabilities. The latter study is based on an analysis of four different NCES surveys, which were used to address enrollment in postsecondary education, access to postsecondary education, persistence to degree attainment, and early labor market outcomes and graduate school enrollment rates of college graduates with disabilities. (See *Students with Disabilities in Postsecondary Education: A Profile of Preparation, Participation, and Outcomes*, NCES 1999–187, by L. Horn and J. Berktold. Washington, DC: 1998.)

The two *Surveys on Distance Education Courses Offered by Higher Education Institutions*, conducted first in late 1995, and again during winter 1998–99, were the first to collect nationally representative data about distance education course offerings in higher education institutions. The two studies differed in their samples and variations in question wording. Further, data from the 1995 study was not imputed for item nonresponse. However, comparisons between the two studies are possible when using the subset of higher education institutions from the 1998–99 study.

The 1995 *Survey on Remedial Education in Higher Education Institutions* was conducted to provide current national estimates on the extent of remediation on college campuses. Results from this survey update the information collected in two earlier NCES surveys for academic years 1983–84 and 1989–90; because PEQIS was not in existence at those times, these surveys were conducted under FRSS. (See section 1 of this chapter.) In addition, although the 1995 survey was not designed as a comparative study, the survey results can be compared with data from the IPEDS Institutional Characteristics Survey: PEQIS estimated that 78 percent of institutions offered at least one remedial course for freshmen in fall 1995, and IPEDS estimated that 79 percent of institutions offered remedial courses in academic year 1993–94. Results from this PEQIS survey can be compared at the student level with institutional surveys conducted by the American Council on Education and an earlier study by the

Southern Regional Education Board. However, these studies asked about freshmen *needing* remediation rather than about freshmen *enrolled* in remedial courses. Remedial enrollments can also be examined from postsecondary transcripts collected during the National Longitudinal Study of the High School Class of 1972 and the High School and Beyond/Sophomores Study. (See chapters 7 and 8.) Institutional reports of remedial enrollments in all of these surveys are substantially higher than student self-reports collected in the NCES National Postsecondary Student Aid Study (NPSAS). (See chapter 16.)

The *Survey on Deaf and Hard of Hearing Students in Postsecondary Education* was conducted in 1993. Comparisons of the estimate of deaf and hard of hearing students obtained from this PEQIS survey with estimates from other surveys show considerable variation due to differences in methodologies and populations of interest. Because the PEQIS study was not designed as a comparative study, the precise reasons for the differences in the estimates from the various sources cannot be answered with the available data. The PEQIS estimate of 20,040 deaf and hard of hearing students in 1992–93 is much lower than the 258,197 national estimate of students with hearing impairments based on student self-reports in the 1989–90 NPSAS. However, the estimate from an earlier institutional study conducted by Gallaudet College (now University) is more in line with the PEQIS estimate—10,400 hearing impaired students enrolled in postsecondary institutions in 1978, including the 2,000 students enrolled at Gallaudet and the National Technical Institute for the Deaf (NTID). The NCES estimate for that year, based on institutional data, was 11,256 "acoustically impaired" students enrolled in postsecondary institutions, excluding Gallaudet and NTID.

Contact Information

For content information on PEQIS, contact:

Bernard R. Greene
Phone: (202) 502–7348
E-mail: bernard.greene@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

Methodology and Evaluation Reports

Methodology discussed in technical notes to survey reports. Some recent reports are listed below.

Campus Crime and Security at Postsecondary Education Institutions, NCES 97–402, by L. Lewis and E. Farris. Washington, DC: 1997.

Distance Education at Postsecondary Education Institutions: 1997–98, NCES 2000–013, by L. Lewis, K. Snow, E. Farris, and D. Levin. Washington, DC: 2000.

Distance Education in Higher Education Institutions, NCES 98–062, by L. Lewis, D. Alexander, and E. Farris. Washington, DC: 1998.

Features of Occupational Programs at the Secondary and Postsecondary Education Levels, NCES 2001–018, by R. Phelps, B. Parsad, E. Farris, and L. Hudson. Washington, DC: 2001.

An Institutional Perspective on Students with Disabilities in Postsecondary Education, NCES 1999–046, by L. Lewis and E. Farris. Washington, DC: 1999.

Chapter 28: Other NCES Surveys and Studies

The final chapter of the Handbook covers five additional projects sponsored by NCES.

FIVE MORE NCES SURVEYS AND STUDIES:

- School Crime Supplement
- School Survey on Crime and Safety
- High School Transcript Studies
- Library Cooperatives Survey
- IEA Civics Study

1. SCHOOL CRIME SUPPLEMENT (SCS)

Overview

The School Crime Supplement (SCS) is conducted periodically as an enhancement to the National Crime Victimization Survey (NCVS), which is administered by the Bureau of Justice Statistics (BJS), U.S. Department of Justice. The NCVS is an ongoing household survey that gathers information on the criminal victimization of household members age 12 and older. NCES and BJS jointly designed the SCS for the purpose of studying the relationship between victimization at school and the school environment.

The SCS gathers data on nationally representative samples of approximately 10,000 students who are between the ages of 12 and 18 and who have attended school at some point during the 6 months preceding the interview. Only crimes that occurred at school during this 6-month period are covered. Topics include victimization in school, avoidance behaviors, weapons, gangs, availability of drugs and alcohol in school, and preventive measures employed by the school. The SCS was fielded in 1989, 1995, 1999, and 2001. Future administrations are planned at 2-year intervals.

Sample Design

Survey estimates for the NCVS are derived from a stratified, multistage cluster sample. The primary sampling units (PSUs) composing the first stage of the sample are counties, groups of counties, or large metropolitan areas. Large PSUs are included in the sample automatically and are considered to be self-representing since all of them are selected. The remaining PSUs (called nonself-representing because only a subset of them is selected) are combined into strata by grouping PSUs with similar geographic and demographic characteristics, as determined by the decennial census.

The households for the NCVS sample are drawn according to the sample design based on the decennial census. The two remaining stages of sampling are designed to ensure a self-weighting probability sample of housing units and group-quarter dwellings within each of the selected areas. (Self-weighting means that, prior to any weighting adjustments, each sample housing unit had the same overall probability of being selected.) This involves a systematic selection of enumeration districts, with a probability of selection proportionate to their population size, followed by the selection of segments (clusters of approximately four housing units each) from within each enumeration district. To account for units built within each of the sample areas after the decennial census, a sample of permits issued for the construction of residential housing is drawn. Jurisdictions that do not issue building permits are sampled using small land-area segments. These supplementary procedures, though yielding a relatively small portion of the total sample, enable persons living in housing units built after the decennial census to be properly represented. Approximately 43,000 housing units and other living quarters were designated for the 1999 NCVS sample.

In order to conduct field interviews for the NCVS, the sample is divided into six groups, or rotations. Each group of households is interviewed seven times—once every 6 months over a period of 3 years. The initial interview is used to bound the interviews (establishing a timeframe to avoid duplication of crimes on subsequent interviews), but is not used to compute the annual estimates. Each rotation group is further divided into six panels. A different panel of households, corresponding to one-sixth of each rotation group, is interviewed each month during the 6-month period. Because the NCVS is continuous, newly constructed housing units are selected as described above, and assigned to rotation groups and panels for subsequent incorporation into the sample. A new rotation group enters the sample every 6 months, replacing a group phased out after 3 years.

All age-eligible individuals in a selected household become part of the panel. NCVS interviews are conducted with each household member who is 12 years old or older. Once all NCVS interviews are completed, an SCS interview is given to household members who were enrolled in primary or secondary education programs leading to a high school diploma sometime during the 6 months prior to the interview. For the 1989 and 1995 SCS, 19-year-old household members were considered eligible for the SCS interview. The upper age range was lowered to 18 for eligibility in the 1999 SCS. Home-schooled students are not surveyed.

Data Collection and Processing

The SCS questionnaire is designed to record the incidence of crime and criminal activity occurring inside a school, on school grounds, or on a school bus during the 6 months preceding the interview. Two modes of data collection were used through the 1999 SCS: paper-and-pencil interviewing (PAPI), which can be conducted in person or over the phone, and computer-assisted telephone interviewing (CATI). For 2001, the CATI questionnaire was replaced by an instrument coded using computer-assisted survey execution system (CASES) software. Interviews are conducted with the subject student between January and June; one-sixth of the sample is covered each month. There were 8,398 SCS interviews completed in 1999, 9,954 in 1995, and 10,449 in 1989. The U.S. Bureau of the Census collects the data. Interviewers are instructed to conduct interviews in privacy unless respondents specifically agree to permit others to be present. Most interviews are conducted over the telephone, and most questions require "yes" or "no" answers, thereby affording respondents a further measure of privacy. While efforts are made to assure that interviews about student experiences at school are conducted with the students themselves, interviews with proxy respondents are accepted under certain circumstances. These include interviews scheduled with a child between the ages of 12 and 13 where parents refuse to allow an interview with the child; interviews where the subject child is unavailable during the period of data collection; and interviews where the child is physically or emotionally unable to answer for him/herself.

Weighting

Weighting compensates for differential probabilities of selection and nonresponse. The NCVS weights are a combination of household-level and person-level adjustment factors. Adjustments are made to account for nonresponse at both levels. Next, additional factors are applied to reduce the variance of the estimate by correcting for differences between the sample distribution of age, race, and sex, and known population distributions of these characteristics. The resulting weights are assigned to all interviewed households and persons on the file. A special weighting adjustment is then made for the SCS respondents. Noninterview adjustment factors are computed to adjust for SCS interview nonresponse. Finally, this noninterview factor is applied to the NCVS person-level weight for each SCS respondent.

Imputation

Because item response rates are high (in all administrations, rates were mostly over 95 percent of all eligible respondents), no imputation is performed.

Sampling Error

To adjust the standard errors to account for the SCS sample design, the Census Bureau developed three generalized variance function (GVF) constant parameters. The GVF represents the curve fitted to the individual standard errors that are calculated using the jackknife repeated replication technique. For the 1989 and 1995 SCS surveys, the three constant parameters (a, b, and c) derived from the curve-fitting process were:

Year	a	b	c
1989	0.00001559	3,108	0.000
1995	-0.00006269	2,278	1.804
1999	-0.00026646	2,579	2.826

To adjust the standard errors associated with percentages, the following formula is used:

$$\text{standard error of p} = \sqrt{\frac{bp(1.0-p)}{y} + \frac{cp(\sqrt{p}-p)}{\sqrt{y}}}$$

where p is the percentage of interest expressed as a proportion and y is the size of the population to which the percentage applies. The estimated standard error of the proportion is then multiplied by 100 to make it applicable to the percentage.

To calculate the adjusted standard errors associated with population counts, the following applies:

$$\text{standard error of x} = \sqrt{ax^2 + bx + cx^{3/2}}$$

where x is the estimated number of students who experienced a given event (e.g., violent victimization).

Coverage Error

The decennial census is used for sampling housing units in the NCVS. To account for units built since the census was taken, supplemental procedures are implemented. (See earlier section on Sample Design.) Coverage error in the NCVS (and SCS), if any, would result from coverage error in the census and the supplemental procedures.

Unit Nonresponse

Because interviews with students can only be completed after households have responded to the NCVS, the unit completion rate for the SCS reflects both the household interview completion rate and the student interview completion rate. The household completion rates were 93.8 percent in 1999, 95.1 percent in 1994, and 96.5 percent in 1989. The student completion rates were 77.6 percent in 1999, 77.5 percent in 1995, and 86.5 percent in 1989. Multiplying the household completion rate by the student completion rate produced an overall SCS response rate of 72.9 percent in 1999, 73.7 percent in 1995, and 83.5 percent in 1989.

Item Nonresponse

Item response rates for the SCS have been high. In all administrations, most items were answered by over 95 percent of all eligible respondents. The only exception was the household income question, which was answered by approximately 86.0 percent of all households in 1999 and approximately 90.0 percent of all households for both 1995 and 1989. Due to their sensitive nature, income and income-related questions typically have relatively lower response rates than other items.

Measurement Error

Measurement error can result from respondents' different understandings of what constitutes a crime, memory lapses, and reluctance or refusal to report incidences of victimization. A change in the screener procedure between 1989 and 1995 probably resulted in the reporting of more incidences of victimization and more detail on the types of crime (and presumably more accurate data) in 1995 than in 1989. (See Data Comparability below for further explanation.) Differences in the questions asked in the NCVS and SCS, as well as the sequencing of questions (SCS after NCVS), might lead to better recall in the SCS. (See below.)

Data Comparability

Respondents to the SCS are asked two separate sets of questions regarding personal victimization. The first set of questions is part of the NCVS, and the second set is part of the SCS. The following have an impact on the comparability of data on victimization: (1) differences between the 1989 and 1995 victimization items on the NCVS; and (2) differences between SCS items and NCVS items for collecting similar data.

Differences between 1989 and 1995 and later NCVS Victimization Items. The NCVS questions capture data on up to six separate incidents of victimization reported by respondents. These questions cover several different dimensions of victimization, including the nature of each incident, where it occurred, what losses resulted, and so forth. Changes to the NCVS screening procedure put in place in 1992 make cross-year comparisons difficult. The victimization screening procedure used in 1995 and later years was meant to elicit a more complete tally of victimization incidents than the one used in 1989. For instance, it specifically asked whether respondents had been raped or otherwise sexually assaulted, whereas the 1989 screener did not. Therefore, cross-year changes in reported victimization rates based on NCVS items may only be the result of changes in how questions were asked and not of actual changes in the incidence of victimization. Refer to the BJS report, *Effects of the Redesign of Victimization Estimates*, for more details on this issue. (See Methodology and Evaluation Reports at the end of this section.)

Because NCVS questionnaires are completed before students are given the SCS questionnaires, it is likely that the changes to the NCVS screening procedure differentially affected responses to the 1989 and the 1995 and later SCS victimization items. Although it is not possible to test this assumption, it is nevertheless reasonable to expect that the more detailed victimization screening instrument led to better victimization recall by SCS respondents in later years than in 1989.

Differences between 1995 and 1999 NCVS and SCS Items. The SCS asks a less detailed set of victimization questions than are asked in the NCVS. Because these questions were not modified between 1989 and 1995, they are more generally comparable for the 2 years. However, the SCS victimization questions were changed in 1999 to specifically ask respondents only to provide information about incidents not previously reported in the main NCVS questionnaire. Thus, unlike prior SCS analyses, in 1999 the prevalence of victimization was calculated by including incidents reports by students on both the NCVS and SCS portions of the instrument.

Additional changes were made in the 1999 SCS. Prior to this year, in 1989 and 1995, students were asked only how easy or hard it was to obtain alcohol or particular drugs at school. In 1999, for the first time, students were asked about alcohol or drugs at school in two parts. There were first asked whether it was possible to obtain alcohol or certain drugs at school. If it was possible to obtain alcohol or a certain drug, they were then asked about the degree of difficulty in obtaining it. Moreover, in 1999, the SCS reworded questions about respondents bring weapons to school. Specifically, students were asked about only guns and knives in the 1999 SCS, while the 1995 SCS asked about other types of weapons as well. The 1999 SCS also covered topics not previously included, such as the use of hate words, the presence of hate-related graffiti, and the prevalence of bullying at school.

Comparisons with Other Related Survey. NCVS/SCS data have been analyzed and reported in conjunction with several other surveys on crime, safety, and risk behaviors. (See *Indicators of School Crime and Safety, 1998*, listed in the Methodology and Evaluation Reports section below.) These other surveys include three NCES surveys: the School Safety and Discipline component of the 1993 National Household Education Survey; teacher victimization items on the Teacher Questionnaire component of the 1993–94 Schools and Staffing Survey; and the Fast Response Survey System's Principal/School Disciplinarian Survey, conducted periodically. Other related surveys and studies include the National School-Based Youth Risk Behavior Survey (YRBS), an epidemiological surveillance system developed by the Centers for Disease Control and Prevention to monitor the prevalence of youth behaviors that most influence health; the School Associated Violent Death Study (SAVD), an epidemiological study developed by the Centers for Disease Control and Prevention in conjunction with the Departments of Education and Justice to describe the epidemiology of school-associated violent death in the United States and identify potential risk factors for these deaths; and Monitoring the Future, an annual ongoing survey conducted by the University of Michigan's Institute for Social Research to study changes in important values, behaviors, and lifestyle orientations of contemporary American youth.

Readers should exercise caution when doing cross-survey analyses using these data. While some of the data were collected from universe surveys, most were collected from sample surveys. Also, some questions may appear the same across surveys when, in fact, they were asked of different populations of students, in different years, at different locations, and about experiences that occurred within different periods of time. Because of these variations in collection procedures, timing, phrasing of questions, and so forth, the results from the different sources are not strictly comparable.

Contact Information

For content information on SCS, contact:

NCES

Kathryn Chandler
Phone: (202) 502–7486
E-mail: kathryn.chandler@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

BJS

Michael Rand
Phone: (202) 616–3494
E-mail: randm@ojp.usdoj.gov

Methodology and Evaluation Reports

The references listed below were either published by the U.S. Department of Education, National Center for Education Statistics (indicated by an NCES number), or published by the U.S. Department of Justice, Bureau of Justice Statistics. See technical notes for discussion of methodology.

General

U.S. Department of Justice, Bureau of Justice Statistics. *Criminal Victimization in the United States, 1994,* NCJ–162126. Washington, DC: 1997.

Uses of Data

Indicators of School Crime and Safety, 2000, NCES 2001–017, by Phillip Kaufman, Xianglei Chen, Susan P. Choy, Sally A. Ruddy, Amanda K. Miller, Jill K. Fleury, Kathryn A. Chandler, Michael R. Rand, Patsy Klaus, and Michael G. Planty. Washington, DC: 2000.

Survey Design

U.S. Department of Justice, Bureau of Justice Statistics. *Effects of the Redesign on Victimization Estimates,* NCJ–164381, by C. Kindermann, J. Lynch, and D. Cantor. Washington, DC: 1997.

2. SCHOOL SURVEY ON CRIME AND SAFETY (SSOCS)

Overview

The School Survey on Crime and Safety (SSOCS) was inaugurated in 2000. By collecting information from school principals in U.S. elementary and secondary schools, it provides detailed information on school crime and safety from the schools' perspective. Measuring the extent of school crime is important for many reasons. The safety of students and teachers is a primary concern, but the nature and frequency of school crime have other important implications as well. Safety and discipline are necessary for effective education. In order to learn, students need a secure environment where they can concentrate on their studies. Further, school crime affects school resources, sometimes diverting funds from academic programs or decreasing schools' ability to attract and retain qualified teachers.

Despite the need for information about school crime, most of the data about it are limited and anecdotal in nature. Schools and policymakers have difficulty knowing which media reports reflect problems that are nationwide and which are relevant only to some schools. Schools also need to know how they compare to other schools nationwide in their policies and programs. For example, there might appear to be a trend toward certain types of school policies (e.g., metal detectors), yet there is often little information about the prevalence of such policies. SSOCS addresses this need by collecting nationally representative data and providing measures of change over time.

Uses of Data

SSOCS is currently NCES' primary source of school-level data on crime and safety. Some of the topics that may be examined are the following:

- frequency and types of crimes at schools, including homicide, rape, sexual battery, attacks with or without weapons, robbery, theft, and vandalism;
- frequency and types of disciplinary actions such as expulsions, transfers, and suspensions for selected offenses;
- perceptions of other disciplinary problems such as bullying, verbal abuse, and disorder in the classroom;
- description of school policies and programs concerning crime and safety; and
- description of the pervasiveness of student and teacher involvement in efforts that are intended to prevent or reduce school violence.

The survey data also support analyses of how these topics are related to each other, and how they are related to various school characteristics.

Sample Design

The SSOCS is a nationally representative cross-sectional survey of about 3,000 public elementary and secondary schools. The sampling frame for the 2000 SSOCS was constructed from the public school universe file created for the 1999–2000 Schools and Staffing Survey (SASS). Only "regular" schools (i.e., excluding schools in the outlying U.S. territories, ungraded schools, and those with a high grade of kindergarten or lower) are eligible for SSOCS.

The sample is first allocated to three instructional levels: elementary schools, middle schools, and secondary/combined schools. Within each instructional level, the sample is further allocated to substrata defined by type of locale, size class, and minority status.

SSOCS was first administered in 2000. It will next be administered in 2003–04, and then NCES plans to conduct SSOCS every 2 years in order to provide continued updates on crime and safety in U.S. schools.

Data Collection and Processing

SSOCS is a mail survey with telephone follow up. The questionnaire is mailed to the school principal. Telephone prompts begin approximately 10 days after the mailout. Fax submissions are accepted.

Returned questionnaires are examined for quality and completeness using both visual and computerized edits. Depending on the total number of items that have missing or problematic data, and on whether those items have been designated as key data items, data quality issues are resolved by recontacting the respondents or by imputation. Westat is the contractor for SSOCS.

Weighting

The SSOCS base weight is the reciprocal of the probability of selecting a school for the sample. To calculate unit nonresponse, adjustment factors are calculated within selected weighting classes, and these factors are applied to the base weights.

Imputation

NCES plans to impute for item nonresponse.

Sampling Error

Standard errors of the estimates are estimated using a jackknife replication method. The estimated standard errors are computed using WesVar.

Future Plans

The next administration will be in 2003–04.

Contact Information

For content information on SSOCS, contact:

Kathryn Chandler
Phone: (202) 502–7486
E-mail: kathryn.chandler@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

Methodology and Evaluation Reports

No documentation has been published as of February 2003.

3. HIGH SCHOOL TRANSCRIPT (HST) STUDIES

Overview

The value of school transcripts as objective, reliable measures of crucial aspects of students' educational experiences is widely recognized. With respect to level of detail, accuracy, and completeness, transcript data are superior to student self-reports of exposure to learning situations. Transcript studies inform researchers and policymakers about the coursetaking patterns of students, which can then be analyzed in relation to the students' academic performance on assessment tests. Since 1982, NCES has conducted six high school transcript studies.

The 1982 study was part of the first follow up to the High School and Beyond (HS&B) Study. (See chapter 8.) Transcripts were collected for members of the 1980 HS&B sophomore cohort who were seniors in 1982. Another transcript study was conducted in conjunction with the 1992 second follow up to the National Education Longitudinal Study of 1988 (NELS:88). (See chapter 6.) Four transcript studies are associated with the National Assessment of Educational Progress (NAEP). (See chapter 20.) Results from the 1987 High School Transcript Study (from schools selected for the 1986 NAEP) were used to compare coursetaking patterns of 12th-grade students in 1982 and 1987. The 1990 HST study, conducted in conjunction with the 1990 NAEP, tracked changes in the curricular patterns of high school students since 1987. The 1994 and 1998 HST studies were conducted in conjunction with those years' NAEP collections. These studies further monitor students' coursetaking behavior.

Sample Design

Sample design is essentially similar across the various administrations of the HST studies: multistage, stratified, and clustered design. However, there are differential rates of oversampling among the studies to reflect special interests. For instance, the 1987 study oversampled students with disabilities and the 1994 and 1998 studies oversampled minority students. Design differences are noted below and in the later section on Data Comparability. The transcript studies are grouped according to the major NCES survey with which they are associated.

The 1998, 1994, 1990, and 1987 Transcript Studies (conducted in conjunction with NAEP). The NAEP Transcript Studies were conducted using nearly identical methodologies and techniques.

The 1998 High School Transcript Study: The 1998 HST sample is nationally representative at both the school and student levels. The sample was comprised of schools selected for the NAEP main sample that had 12th-grade classes and were within the 58 PSUs selected for the HST study. A subsample of 322 schools was selected from the eligible NAEP sample, consisting of 269 public schools and 53 nonpublic schools. In order to maintain as many links as possible with NAEP scores, replacement schools that were used in NAEP were also asked to participate in the transcript study, as opposed to sampling the NAEP refusal schools. Of the 322 schools in the original sample, 264 participated, of which 232 cooperated with both NAEP and HST and maintained links between students' transcript and NAEP data.

A total of 28,764 students were selected for inclusion in the HST study. Of these, 27,183 students were from schools that maintained their NAEP administration schedules and were identified by their NAEP booklet numbers. Another 500 students were from schools that participated in NAEP but had lost the link between student names and NAEP booklet numbers, and 1,081 were from schools that did not participate in NAEP. Of the 28,764 students in the original sample, 25,248 were deemed eligible for the transcript study, and 24,218 transcripts were collected and processed.

The 1994 High School Transcript Study. The 1994 HST sample of schools was nationally representative of all high schools in the United States. A subsample of 333 public schools and 47 private schools were drawn from the lists of eligible NAEP public and private schools. One of these schools had no 12th-grade students, and was not included in the HST study. Of the 379 remaining schools, 340 participated in the 1994 HST study. The student sample was representative of graduating seniors from each school. Only those students were included whose transcripts indicated that they had graduated between January 1, 1994, and November 21, 1994. Approximately 90 percent of students in the 1994 HST study also participated in the 1994 NAEP. The remaining students were sampled specifically for the transcript study, either because their schools did not agree to participate in the 1994 NAEP or because the schools participated in the NAEP study but did not retain the lists linking NAEP IDs to student names. The 1994 HST study also included special education students who were excluded from the 1994 NAEP. High school transcripts were collected for 25,494 from an eligible sample of 26,045 students.

The 1990 High School Transcript Study. The sample of schools was nationally representative of schools with grade 12 or having 17-year-old students. (Some 370 schools were selected for the sample; 8 of these had no 12th-grade students.) The sample of students was representative of graduating seniors from each school. These students attended 330 schools that had previously been sampled for the 1990 NAEP. Approximately three-fourths of the sampled students had participated in the 1990 NAEP assessments. The remaining students attended schools that did not participate in the NAEP or did not retain the lists linking student names to NAEP IDs. As with the 1994 HST study, only schools with a 12th grade were included, and only students who graduated from high school in 1990 were included. The 1990 HST study also included special education students who had been excluded from the 1990 NAEP. In spring 1991, transcripts were requested for 23,270 students who graduated from high school in 1990; 21,607 transcripts were received.

The 1987 High School Transcript Study. The schools in the 1987 HST study were a nationally representative sample of 497 secondary schools that had been selected for the 1986 NAEP assessments. The 1987 HST student sample represented an augmented sample of 1986 NAEP participants who were enrolled in the 11th grade and/or were 17 years old in 1985–86 and who successfully completed their graduation requirements prior to fall 1987. The HST study included (1) students who were selected and retained for the 1986 NAEP assessment; (2) students who were sampled for the 1986 NAEP but were deliberately excluded due to severe mental, physical, or linguistic barriers; and (3) all students with disabilities attending schools selected for the 1986 assessment. Four of the participating schools had no eligible students without disabilities. Of the 497 schools selected for the HST study, 433 participated in the study. There were 35,180 graduates in the sample, for whom 34,140 transcripts were received.

Westat, Inc. conducted the NAEP HST studies.

The 1992 High School Transcript Study. This transcript study was conducted as part of the NELS:88 second follow up—see chapter 6. A total of 2,258 schools were identified in the second follow-up tracing of the NELS:88 first follow-up sample. Since the HST component was limited to 1,500 schools, it was necessary to select a

294 BEST COPY AVAILABLE

sample of schools. All schools identified as having four or more first follow-up sample members enrolled were included in the school-level sample with certainty (1,030 schools, probability = 1.0), and random samples were selected for retention from schools identified as having three first follow-up members (45 out of 60 schools, probability = 0.75), two first follow-up members (104 out of 160 schools, probability = 0.65), and one first follow-up member (321 out of 1,008, probability = 0.31845). (Note that by the time of data collection, only 1,374 of the 1,500 schools contained at least one NELS sample member.) Transcript data were requested for all students in the 1,374 selected schools.

In addition, transcripts were collected for all dropouts, early graduates, and 12th-grade sample members ineligible for the base year, first follow-up, and second follow-up surveys owing to a language, physical, or mental barrier (triple ineligibles). Including triple ineligibles improved comparability with the 1987 and 1990 NAEP-based transcript studies, which included special education students excluded from NAEP administrations as well as NAEP-eligible students. This added 468 schools to the sample.

Of the 1,842 schools in the 1992 sample, 1,543 participated in the 1992 study. Transcripts were requested for 19,320 students, and 17,285 transcripts were received. This study was conducted by the National Opinion Research Center (NORC) at the University of Chicago.

The 1982 High School and Beyond (HS&B) Transcript Study. The first transcript study was a component of the HS&B first follow up. The 1982 study included 1,899 secondary schools—999 HS&B sampled schools and 900 schools to which students selected for the transcript survey had transferred (and for which no data collection activities other than transcript collection were carried out). Of these 1,899 schools, 1,720 provided transcripts. The total student sample size was 18,427 students. From among the 1980 sophomores selected for the HS&B first follow up, 12,309 cases were retained in the HST sample with certainty—12,034 cases in the probability sample plus 275 nonsampled co-twins. In addition, a systematic sample of 6,118 cases was subsampled from the 17,703 remaining first follow-up selections, with a uniform probability of approximately .35. Transcripts were collected for 15,941 of the 18,427 students. The NORC at the University of Chicago conducted this study.

Data Collection and Processing

The procedures for transcript and other data collection and processing are similar for the various HST studies. The description in this section pertains mostly to the five NAEP-based transcript studies. The 1998 HST procedures illustrate the process.

NAEP field workers requested sample materials for the 1998 HST study when they first went to a school as part of the 1998 NAEP, and they collected these materials when they returned to the school for sampling. The sample materials included a list of courses offered for each of four consecutive years from 1994 to 1998; a completed School Information Form (SIF); and three transcripts of students who graduated in 1998 (representing a "regular" student, one with honors courses, and one with special education courses). An SD/LEP questionnaire was completed for students with a disability or with limited English proficiency by the person most knowledgeable about the student. The School Questionnaire—a 54-item questionnaire that asked for information about school, teacher, and home factors that might relate to student achievement—was completed by a school official (usually the principal) as part of NAEP.

The SIF requested information about the school in general, sources of information within the school, course description materials, graduation requirements, grading practices, and the format of the school transcripts or as part of the HST data collection process for non-NAEP participating schools.

In schools that did not participate in NAEP, the field worker first selected a sample of students, then requested transcripts for those students and followed the procedures for NAEP participants for reviewing and shipping transcripts. The SIF was also completed and course catalogs for the past four academic years were collected. The information in the catalogs was documented by completing the Course Catalog Checklist. At this point the procedure was different. Rather than obtaining and annotating three example transcripts, the field worker used the Transcript Format Checklist to annotate three actual transcripts from among those that were collected.

In the non-NAEP participating schools, the process of generating a sample of students began when the school produced a listing of all students who graduated from the 12th grade during the spring or summer of 1998. This list was requested during the preliminary call placed to the school when it was determined that the school would participate in the HST. The following information was collected for each student in the HST: exit status; sex; date of birth (month/year); race/ethnicity; whether the student had a disability (SD); whether the student was classified as Limited English Proficiency (LEP); whether the student was receiving Title I services; and whether the student was a participant in the National School Lunch Program. These data were collected either with the list of 1998 graduates or after sampling, depending on which procedure was easier for the school. SD/LEP questionnaires were not collected for students in schools that had not participated in NAEP.

Each of the courses entered on the transcripts were coded using a common course-coding system, a modification of the *Classification of Secondary School Courses* (CSSC). The CSSC—which contains approximately 2,000 course codes—is a modification of the Classification of Instructional Programs (CIP) used for classifying college courses. Both systems use a three-level, six-digit system for classifying courses. The CSSC uses the same first two levels as the CIP, represented by the first four digits of each code. The third level of the CSSC (the fifth and sixth digits of the course code) is unique to the CSSC and represents specific high school courses.

For all NAEP transcript studies, courses appearing on student transcripts were also coded to indicate whether they were transfer courses, held off campus, honors or above grade-level courses, remedial or below grade-level courses, or designed for students with Limited English Proficiency and/or taught in a language other than English.

Credit and grade information reported on transcripts also needed to be standardized. Standardization of credit information was based on the Carnegie Unit, defined as the number of credits a student received for a course taken every day, one period per day, for a full school year. (Note that the 1982 High School and Beyond Transcript Study provided course totals rather than Carnegie Units.) Coders converted numeric grades to standardized letter grades unless the school documents specified other letter grade equivalents for numeric grades.

The Computer Assisted Coding and Editing (CACE) system was designed specifically for coding high school catalogs. CACE has two major components: (1) a component for selecting and entering the most appropriate CSSC code and "flags" for each course in a catalog; and (2) a component for matching each entry appearing on a transcript with the appropriate course title in the corresponding school's list of course offerings.

Each stage of the data coding and entering process included measures to assure the quality and consistency of data. Measures to maintain the quality of data entry on transcripts included: 100 percent verification of data entry; review of all transcripts where the number of credits reported for a given year (or the total number of credits) was not indicative of the school's normal course load or graduation requirements; and reconciliation of transcript IDs with the list of HST-valid IDs. Catalog coding reliability was maintained by conducting reliability checks. At least 10 percent of each school's course offerings were re-entered by an experienced coder and the results compared with those of the original coder. If less than 90 percent of the entries agreed, the catalog was completely reviewed and any necessary changes were made. Agreement of 90 percent or better was found for approximately 85 percent of the school catalogs during the first review.

An additional quality check took place when the CACE files for a school were converted to delivery format. Reports listing frequencies of occurrences that might indicate errors were sent to the curriculum specialist for review. Each file was then assigned a status of 1 for complete, 2 for errors in transcript entry, 3 for errors in catalog coding and associations, or 4 for computer errors. A file with a status of 2, 3, or 4 was returned to Computer Assisted Data Entry (CADE) and CACE for correction, a new report was generated, and the report was again reviewed. This process was repeated until the file had a status of 1, indicating that it was complete and correct.

Weighting

The sampling weights for the HST studies are designed primarily to represent differential sampling and response rates. Only the 1998 procedures are described below. (For details on weighting in the other studies, see the relevant technical manuals.)

Two types of weights were created in the 1998 HST:

- HST base weights for all students who participated in the 1998 HST study—that is, for whom a transcript was received and coded; and
- HST-NAEP linked weights for students who participated in both the 1998 HST and the 1998 NAEP. Linked weights were computed separately for writing, 25-minute reading, 50-minute reading, civics, and civics trend assessment students. Each assessment sample represents the full population, so each of the five sets of assessment-linked weights aggregate separately to the population totals.

In each set of weights, the final weight attached to an individual student record reflected two major aspects of the sample design and the population surveyed. The first component, the base weight, reflected the probability of selection in the sample (the product of the probability of selecting the primary sampling unit, the probability of selecting the school within the primary sampling unit, and the probability of selecting the student within the school). The second component resulted from the adjustment of the base weight to account for nonresponse within the sample and to ensure that the resulting survey estimates of certain characteristics (race/ethnicity, size of community, and region) conformed to those known reliably from external sources.

The final *HST student weights* were constructed in five steps:

(1) The student base weights (or design unbiased weight) were constructed as the reciprocal of the overall probability of selection.

(2) School nonresponse factors were computed, adjusting for schools that did not participate in the HST study. For the linked weights, adjustment factors were assigned for each session type (writing/civics, reading, and civics trend). The school nonresponse factors for the linked weights were also slightly different than the corresponding HST student weight school nonresponse factors, to account for schools that refused to participate in NAEP.

(3) Student nonresponse factors were computed, adjusting the weights of responding students to account for nonresponding students. Definitions of responding and nonresponding students differed for the HST weights and the linked weights.

(4) Student trimming factors were generated to reduce the mean squared error of the resulting estimates. Another purpose of the trimming was to protect against a small number of large weights from dominating the resulting estimates of small domains of interest.

(5) The final step was poststratification, the process of adjusting weights proportionally so that they aggregate within certain subpopulations to independent estimates of these subpopulation totals. These independent estimates were obtained from the Current Population Survey (CPS) estimates for various student subgroups. As the CPS estimate has smaller sampling error associated with it, this adjustment should improve the quality of the weights.

The *linked student weights* were constructed in a parallel manner, with some differences (e.g., the student base weight incorporated a factor for assignment to NAEP assessments). The school nonresponse factors were also slightly different for the linked weights to account for schools that refused to participate in the NAEP assessments. In addition, there was an extra nonresponse factor computed for the linked weights to adjust for students whose transcripts were included in the HST study but who were absent from (or refused to participate in) a NAEP assessment. The trimming and poststratification steps for the linked weights were similar to those of the HST weights, with some differences. The missing transcript adjustments for the linked weights were very similar to those computed for HST weights.

Imputation

In the 1994 and 1998 HST, for a small percentage of graduated students it was not possible to obtain a transcript. In addition, some transcripts were considered unusable, since the number of standardized credits shown on the transcript was less than the number of credits required to graduate by the school. An adjustment is necessary in the weights of graduated students with transcripts to account for missing and unusable transcripts. To do this adjustment correctly, it is necessary to have the complete set of graduated students, with or without transcripts. Students who did not graduate were not included in this adjustment, but they were retained in the process for poststratification. There are a few students, however, for whom no transcripts were received and the graduation status was unknown. Among these students, a certain percent was imputed as graduating, based on overall percentages of graduating students. The remaining students were imputed as nongraduating. The imputation process was a standard (random within class) hot-deck imputation. For each student with unknown graduation status, a "donor" was randomly selected (without replacement) from the set of all students with known graduation status from the same region, school type, race/ethnicity, age class, school, and sex, in hierarchical order. The two race/ethnicity categories were (1) White, Asian, or Pacific Islander and (2) Black, Hispanic, American Indian, or other. There were two age classes (born before 10/79; born during or after 10/79). Each student with known graduation status in a cell could be used up to three times as a donor for a student in the same cell with unknown graduation status. If insufficient donors were available within the cell, then donors were randomly selected from students in another cell with similar characteristics to the cell in question. A donor had at least to be from the same region, type of school, race category, and age category.

Imputation was done for missing sex data in the 1992 NELS Transcript Study, using the student's first name to determine the sex. In the 1982 HS&B Transcript Study, values were imputed for missing sex and race/ethnicity. Because the 1982 and 1992 studies were part of longitudinal studies covering the same students over time, there were more opportunities to collect information on both sex and race/ethnicity than in the NAEP studies.

Sampling Error

Because of the HST multistage design, jackknife replication was used for variance estimation. In the 1998 HST, a set of 62 replicate weights was attached to each record, one for each replicate. Variance estimation was performed by repeating the estimate procedure 63 times, once using the original full set of sample weights and once each for the set of 62 replicate weights. The variability among replicate estimates was used to derive an approximately unbiased estimate of the sampling variance. This procedure was used to obtain sampling errors for a large number of variables for the whole population and for specified subgroups.

In general, the variability was very small compared to the size of the estimates, although this is not true in cases of infrequently taken courses in the smaller subpopulations. For example, the percentage of White students taking geometry is estimated at 78.08, with a standard error of 1.03 (a ratio of 0.01), while the percentage of Native Americans taking calculus is estimated at 4.14, with a standard error of 1.62 (a ratio of 0.39). (See *The 1998 High School Transcript Study Tabulations*, NCES 2001-498.)

Coverage Error

Potential sources of undercoverage in the HST studies include: (1) incomplete sampling frame data, as no national listing of schools is, or remains for very long, 100 percent complete and accurate (see "Nonsampling Error, Coverage error" in chapters 6, 8, and 20, as relevant to the particular HST study); (2) omissions and errors in school rosters; and (3) deliberate exclusion of certain categories of students—such as students with physical or mental disabilities or non-English speakers, who might find it difficult or impossible to complete demanding cognitive tests and questionnaires. The first two sources are thought to have only a very small impact on HST estimates. The most serious potential source of undercoverage bias for HS&B, NELS, and NAEP studies is believed to be the exclusion of students with physical, mental, or linguistic barriers to assessment or survey participation. While these studies have used similar exclusion criteria for completion of survey forms and testing, specific guidelines differ somewhat across studies, as well as within studies over time. In an effort to minimize the number of exclusions, eligibility criteria were made more specific in 1990.

Because the NAEP and NELS studies collected data on the characteristics of excluded students, undercoverage bias can be quantified. Also, these studies were more inclusive in their transcript components than in their test or questionnaire administration. (See Sample Design above.) It is believed that NAEP transcript studies had no transcript undercoverage due to exclusion of certain students and that the 1992 NELS study had negligible undercoverage of about 2.5 percent for the senior cohort. Although quantifiable exclusion data are not available for the HS&B, given the similarity of eligibility rules in all three studies, it is reasonable to presume that HS&B exclusion rates were between 3 and 6 percent.

Unit Nonresponse

There is unit nonresponse at both the school and student levels in HST studies. In 1998, an unweighted 88 percent of schools participated in the transcript study (compared to 90 percent in the 1994 study, 87 percent in both the 1987 and 1990 studies, 91 percent in the 1982 HS&B study (95 percent for HS&B regular schools vs. 86 percent for transfer schools), and 84 percent in the 1992 NELS study (94 percent for contextual schools vs. 55 percent for noncontextual schools). Response rates, however, varied with characteristics of the sample school. For example, in 1998, despite the high overall response rate, only 71 percent of nonpublic schools responded to the study.

At the student level, transcripts were obtained for 98 percent of eligible students in the 1998 HST study. This rate matches that for the 1994 HST study and is higher than the student-level response rates for the other studies—89 percent in 1992 (92 percent for students in contextual schools versus 74 percent for dropouts and alternative completers); 93 percent in 1990; 97 percent in 1987; and 88 percent in 1982 (89 percent for students in regular HS&B schools versus 72 percent for transfer students).

Item Nonresponse

Rates for item nonresponse have ranged from nonexistent to extremely high, depending on the type of item. As would be expected in transcript studies, course-level items have little if any nonresponse. Specific items include school year, term, and grade in which a course was taken; school-assigned course credits; and standardized course grade. For these items in the 1992 NELS Transcript Study, nonresponse rates ranged from 0 percent for school year to less than 2 percent for school term in which a course was taken. Incompleteness of actual course data, while considered to be limited, is another source of potential bias in a transcript study. Course data may be incomplete for students who transferred from one school to another. Also, it is difficult to assess the completeness of transcript data for dropouts (1982 HS&B and 1992 NELS) because of inconsistencies between enrollment reports of the sample member and the school.

Transcripts often provide other pieces of information useful for analysis of coursetaking patterns: days absent in each school year, class rank, class size, month and year student left school, reason student left school (e.g., dropped out, graduated, transferred), cumulative GPA, participation in specialized courses or programs, and various standardized test scores (e.g., PSAT, SAT, ACT). While nonresponse rates for participation in specialized courses or programs (1.8 percent in 1992) and month/year/reason student left school (less than 4 percent in 1992) are quite low, nonresponse rates for the other items are very high: in 1992, 18 percent nonresponse for class size; 22 percent for cumulative GPA; 23 percent for class rank; 42–44 percent for days absent in each of the 4 high school years; and 67–73 percent for standardized test scores. (Note that although students were asked on a student questionnaire whether and when they planned to take specific tests, some students may not have actually taken the tests; this would explain in part the high nonresponse rates for test scores.) This wide range of item nonresponse rates is comparable to results of the 1982 HS&B Transcript Study and the NAEP transcript studies. For example, the 1982 HS&B study showed 32 percent nonresponse for class rank and class size, 41–47 percent nonresponse for days absent per school year, and 75 percent and above for standardized test scores.

Two key analytic variables are sex and race/ethnicity. Item nonresponse rates for sex have been extremely low, ranging from 0 percent in the 1982 HS&B study and the 1992 NELS study to 0.26 percent in the 1987 NAEP study. For race/ethnicity, nonresponse has ranged from 0 percent in 1982 and 0.7 percent in 1992 to 5.4 percent in 1987.

Measurement Error

Possible sources of measurement error in HST studies are differences between schools and teachers in grading practices (e.g., grade inflation), differences in how data are recorded (although efforts are made to standardize grades and course credits for the HST studies), and errors in keying or processing the transcript data (although the system has many built-in quality checks). The amount of measurement error in any survey or study is difficult to determine, and it is unknown for the HST studies. However, because the transcripts are official school records of students' progress, it is reasonable to presume that there is less measurement error than in other types of data collections, particularly those that are self-reported.

Data Comparability

While there are many similarities among the HST studies conducted thus far, there are also some differences. Users should consider the following:

Sample Design. The overall sample design for the HS&B, NELS:88, and NAEP studies is quite similar. All are large, nationally representative school-based samples that have employed a multistage, stratified, clustered design. However, despite their fundamental similarity, the designs differ somewhat in a number of features. Five differences, in particular, should be considered because of their potential impact on comparative analyses:

Sample sizes. There are differences in sample sizes across the various transcript studies, and marked differences in the distribution of transcript-eligible students across schools. For example, the 1982 HS&B Transcript Study collected 15,941 transcripts from 1,720 schools. In contrast, the 1987 NAEP study collected more than twice as many transcripts (34,140) from a quarter as many schools (433). The 1982 HS&B Transcript Study collected considerably fewer transcripts than were collected in the other transcript studies and from a considerably greater number of schools. This means that comparable estimates across the multiple transcript studies have similar sampling errors despite differences in the total number of transcripts sampled. In fact, sampling errors were often *smaller* for the 1982 estimates. The design effects for years other than 1982 were considerably larger than for 1982, more than offsetting the effects of the larger sample size of transcripts in those other years.

Oversampling. To reflect special interests, different rare student populations and school types have been disproportionately included in the studies. The 1982 HS&B Transcript Study included nonsampled co-twins, and the 1987 NAEP Transcript Study oversampled students with disabilities. The HS&B study oversampled Hispanics; the NELS:88 study oversampled Asians and Hispanics; and the NAEP studies oversampled schools with high percentages of Hispanics and Blacks. All studies oversampled private schools.

Eligibility. While similar exclusion criteria have been used for the main HS&B, NELS:88, and NAEP studies, specific guidelines have differed. Eligibility criteria became more specific in 1990, so comparability between studies should have improved. (See Sample Design above for eligibility criteria for the transcript studies, which have included special education students who were excluded from the main surveys.)

Representativeness of cross-sectional and longitudinal populations. The HS&B and NAEP transcript studies were based on national probability samples of high schools. Although the transcript studies did not always take place in the years the school samples were drawn, the timeframes were close enough to consider the samples a close approximation of a national probability sample of schools for that year. The 1992 NELS transcript study, on the other hand, cannot be considered nationally representative of high schools in 1992. Rather, it represents the schools to which a national probability sample of 8th graders had dispersed 2 and 4 years later.

One fundamental difference among the transcript studies is that the 1982 HS&B study and the 1992 NELS study collected transcripts of students who were still enrolled in school, dropouts, transfers, and GED recipients, whereas the NAEP studies excluded these students. Also, the student samples for the various studies were drawn at different points in students' high school careers so they are not universally representative of the senior classes for the study years. The 1982 HS&B students were sampled when they were sophomores in 1980. Although transferring students were followed to their new schools, the 1982 student sample is not *fully* representative of high school seniors because it does not include (1) eligible students who were not selected in 1980 but who had since transferred into a HS&B school, and (2) 1982 seniors who were not sophomores in 1980. The students for the 1987 NAEP Transcript Study were sampled for the 1986 NAEP when they were juniors and/or 17 years old, but no attempt was made to follow them if they left school as a transfer or dropout, Nor were students who transferred into the school after NAEP sampling included. Additionally, 1987 graduating seniors who were not 1986 juniors had no chance of selection into the study. The 1987 sample, therefore, only approximates the high school graduating class of 1987. The students in the 1990, 1994, and 1998 NAEP studies were sampled in their senior year and were further restricted to seniors who actually graduated in those years. As such, these studies do provide representative samples of each high school's graduates in the respective years. These studies, like the one in 1987, excluded students who transferred out, failed to graduate on time, or who received GEDs. In contrast to these studies, the 1992 NELS in-school samples of students are not necessarily representative of seniors within these schools since they exclude non-NELS 8th graders who may have fed the schools.

Definition of Seniors. *Users should be cautious when comparing data for seniors in a given academic year (e.g., 1991–92) with graduates in a given calendar year (e.g., 1992).* Moreover, not all members of the 1982 HS&B senior cohort and the 1992 NELS senior cohort succeeded in meeting graduation requirements. The transcript data sets generally provide information about both the date and the reason for leaving the school so that the same unit of analysis (e.g., graduates as of a certain point in time) can be determined. (See *Sample Design* differences above.)

Coded Information. In all of these studies, transcripts were obtained from both public and private high schools. Information from these transcripts—including specific courses taken, grades, and credits earned—was coded according to the CSSC coding system and processed into a system of data files designed to be merged with questionnaire and test data files. (See Data Collection above.) In addition to general course information, the CSSC for coding transcript data includes a "disability" flag and a "sequence" flag. The disability flag was added to the CSSC during the 1987 transcript study to indicate whether a course is open to all students or is restricted to disabled students. The sequence flag indicates whether a course is part of a sequence of courses and, if so, its place in that sequence. It was added to the CSSC during the 1990 transcript study.

Unlike the other HST studies, some transcript information was not coded in the 1982 HS&B study. Uncoded information includes the identification of courses as remedial, regular, or advanced; as offered in a different location; or as redesigned for students with disabilities. (The HS&B study also used a different method for identifying students with disabilities than did the other studies.)

As noted above, the HS&B and NELS transcript studies included students who had not yet graduated, who received a GED, who transferred to another school, or who dropped out of school. Transcript information for some of these students is less complete than for seniors who graduated from their sampled school. Dropouts would not necessarily have transcripts spanning the usual 4-year high school career. While attempts were made to obtain transcripts for transferring students, the transfer schools were less cooperative than were schools that were part of the regular school sample.

Contact Information

For content information on the High School Transcript Studies, contact:

1987, 1990, 1994, and 1998 Studies:
Janis Brown
Phone: (202) 502–7419
E-mail: janis.brown@ed.gov

1982 and 1992 Studies:
Jeffrey Owings
Phone: (202) 502–7423
E-mail: jeffrey.owings@ed.gov

Mailing Address:
National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

Methodology and Evaluation Reports

The U.S. Department of Education, National Center for Education Statistics, is the source of the references listed below.

General

High School Transcript Study, 1987, ERIC Document Reproduction Service No. ED315450, by Judy Thorne et al. Washington, DC: 1989.

The 1990 High School Transcript Study Technical Report, ERIC Document Reproduction Service No. ED360375. Washington, DC: 1993.

The 1994 High School Transcript Study Technical Report, NCES 97–262, by Stanley Legum, Nancy Caldwell, Bryan Davis, Jacqueline Haynes, Telford J. Hill, Stephen Litavecz, Lou Rizzo, Keith Rust, Ngoan Vo, and Steven Gorman. Washington, DC: 1997.

1998 Revision of the Secondary School Taxonomy, NCES Working Paper 1999–06, by Denise Bradby and Gary Hoachlander. Washington, DC: 1999.

Procedures Guide for Transcript Studies, NCES Working Paper 1999–05, by Martha Naomi Alt and Denise Bradby. Washington, DC: 1999.

Uses of Data

National Education Longitudinal Study of 1988 Second Follow-Up: Transcript Component Data File User's Manual, NCES 95–377, by Steven J. Ingels, Kathryn L. Dowd, John R. Taylor, Virginia H. Bartot, Martin R. Frankel, Paul A. Pulliam, and Peggy Quinn. Washington, DC: 1995.

The 1990 High School Transcript Study Data File User's Manual, ERIC Document Reproduction Service No. ED361354, by Nancy Caldwell et al. Washington, DC: 1993.

The 1998 High School Transcript Study User's Guide and Technical Report, NCES 2001–477, by Stephen Roey, Nancy Caldwell, Keith Rust, Eyal Blumstein, Tom Krenzke, Stan Legum, Judy Kuhn, Mark Waksberg, Jacqueline Haynes, and Janis Brown. Washington, DC: 2001.

The 1998 High School Transcript Tabulations: Comparative Data on Credits Earned and Demographics for 1998, 1994, 1990, 1987, and 1982 High School Graduates, NCES 2001–498, by Stephen Roey, Nancy Caldwell, Keith Rust, Eyal Blumstein, Tom Krenzke, Stan Legum, Judy Kuhn, Mark Waksberg, Jacqueline Haynes, and Janis Brown. Washington, DC: 2001.

4. LIBRARY COOPERATIVES SURVEY (LCS)

Overview

The Library Cooperatives Survey (LCS) was first administered in 1998 and is scheduled to be conducted at 5-year intervals thereafter. The first survey gathered data for fiscal year (FY) 1997 from about 400 library cooperatives. LCS collects descriptive information about library cooperatives—entities that provide additional services and resources primarily to public, academic, school, and special libraries. Data items include member service measures, such as number of reference transactions and interlibrary loans, training and instruction hours provided to member library staff, and consulting and planning hours. In addition, the library cooperatives report information about membership, size of collection, operating income and expenditures, and staffing.

The survey included 55 data items and covered the following areas: type of organization, geographic area served; whether the general public is directly served;

cooperative membership; operating income; operating expenditures; capital expenditures; cooperative services such as reference, interlibrary loan, training, consulting, Internet access, electronic services, statistics, preservation, union lists, public relations, cooperative purchasing, delivery, advocacy, and outreach programming.

The data from this survey fill a significant gap in library information. The results are extremely useful to federal, state, and local officials in assessing the utility of library cooperatives in sharing resources among various types of libraries. Input elements (e.g., expenditures) can be compared with output elements (e.g., services). Additionally, the FY 97 data serve as a critical baseline for gauging the influence of the 1997 Library Services and Technology Act (LSTA) on library cooperatives. LSTA urges cooperative relationships and resource sharing among various types of libraries.

Data Collection and Processing

The FY 97 data were collected in spring 1998 through a combination of paper forms and electronic forms accessed by respondents via the Internet. A pretest collection of FY 96 data from a sample of approximately 150 respondents in the anticipated universe was conducted in the summer of 1997, using paper forms. At that time, quick-response postcards were mailed to all organizations in the universe (approximately 768) to collect address corrections and qualifying information for the full survey.

Contact Information

For additional information about LCS, contact:

Jeffrey Williams
Phone: (202) 502–7476
E-mail: jeffrey.williams@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

Methodology and Evaluation Reports

Currently, no reports are available for LCS.

5. CIVIC EDUCATION STUDY (CivEd)

Overview

Within the United States there has been growing interest in cross-national comparisons of students' educational achievement. In light of the rapidly changing international political and economic climate, this interest has focused on a concern about the ability of our population to meet the growing challenges of an information society and a desire to maintain our competitive advantage in the world economy. In addition to participation in cross-national comparisons of reading literacy (see chapter 22), adult literacy (see chapter 24), and mathematics and science education (see chapter 21), in 1999 the United States participated in the International Association for the Evaluation of Educational Achievement (IEA) Civic Education Study (CivEd).

Phase I of CivEd began in 1995 and 1996, examining the goals and curriculum of civics education in approximately 20 countries. The product of Phase I, released in 1999, was a volume of case studies describing civics education in participating countries, designed to provide the information needed to develop a framework to guide the construction of an assessment instrument about civic knowledge and behavior. Phase II was the administration of the assessment in the fall of 1999. The assessment measures 9th-grade students' civic knowledge, skills, and attitudes across the following three domains: democracy, national identity and international relations, and social cohesion and diversity.

Components

The 1999 CivEd consisted of three instruments: a student questionnaire, a school questionnaire, and a teacher questionnaire.

Student Questionnaire. The questionnaire contained five types of items: items assessing knowledge of key civic principles and pivotal ideas (civic content items—type 1); items assessing skills in using civic-related knowledge (civic skills items—type 2); items measuring students' concepts of democracy, citizenship, and government (type 3); items measuring attitudes toward civic issues (type 4); and items measuring expected political participation (type 5). Additional survey questions assessed students' perceptions of the climate of the classroom and other background variables. Test questions were multiple-choice.

School Questionnaire. The school questionnaire, completed by the principal, contained questions designed to gather information on the school's general environment, such as size, length of school year, and characteristics of the student body. The school questionnaire also asked questions designed to provide a picture of how civic education is delivered through the school curriculum and school-sponsored activities, as well as the number of staff involved in teaching civic-related subjects.

Teacher Questionnaire. A teacher questionnaire was administered to the teacher of the selected class. However, because the organization of civic education and the role of civic education teachers in U.S. schools differ from those of many other countries in the study, results from the teacher questionnaire were not analyzed in the U.S. report.

Sample Design

The CivEd school sample for the United States was drawn in October 1998, following international requirements as given in the *IEA Civics School Sampling Manual*. The United States sample was a three-stage, stratified, clustered sample. The overall sample design was intended to approximate a self-weighting sample of students as much as possible, with each 9th-grade student in the United States having an approximately equal probability of being selected (within the major school strata).

The first stage included defining geographic primary sampling units (PSUs); classifying the PSUs into strata defined by region and community type; then selecting PSUs with probability proportional to size.

The second stage of sampling was the selection of schools, using a frame developed from two lists. Regular public, Bureau of Indian Affairs, and Department of Defense Education Activity schools were obtained from the 1997 QED list. Catholic and nonpublic schools were obtained from the 1995–96 Private School Survey. (See chapter 3.) Any school having a 9th grade and located within an IEA Civics PSU was included on the school sampling frame. A total of 7,936 schools were on the frame.

The primary variable ordering the schools on the frame was public/private status: a total of 11 private schools and 139 public schools were drawn in the final sample. The measures of size for each school were designed to be proportional to the estimated number of 9th graders in the school, within each implicit stratum. A Keyfitz procedure was carried out to minimize overlap with the 1999 TIMSS-R school sample being fielded in the same PSUs during the same time period. Additionally, for public schools, measures of size were assigned such that those with high minority populations (greater than 15 percent Blacks and Hispanics) had probabilities of selection twice as high as those in the same PSU with low minority populations; for private schools, additional stratification was done by three size groupings, and the two smallest strata were given reduced measures of size to lower the expected sample count of schools in these strata. When drawing the school sample, private schools were ordered first by school type, next by PSU, and last by measure of size. Public schools were ordered first by PSU, next by minority enrollment category, and last by measure of size.

The third stage of sampling was classrooms within schools. Within each participating school, the plan was to randomly select one classroom, preferably in Civics or a related subject, and all students in the classroom were selected. In schools that could not provide a list of classes for grade 9 that (a) included every grade 9 student in the school exactly once, and (b) was preferably a Civics or related class, alternative procedures were used. Classrooms with less than 15 students were collapsed into pseudo-classrooms.

Finally, the teacher of the selected class was asked to complete a questionnaire.

Data Collection and Processing

The CivEd data were collected in fall 1999. States, then school districts, and then schools were contacted about participating in CivEd. Schools were offered an honorarium for their participation and a one-page report indicating how their students did. With these incentives, a school cooperation rate of 89 percent (including substitutes) was secured.

Westat handled the field operations, and hired and trained the external test administrators.

In each school, an original testing session was held, and a makeup session if the student response rate was less than 90 percent. Overall, the student response rate was 92 percent, with only 7 students assessed in makeup sessions. The sessions were administered according to international specifications, and timed as specified in the script and international materials. Most sessions were conducted in the morning with minimal breaks of 3–10 minutes. A total number of 124 schools and 2,811 students participated.

Data were optically scanned.

Weighting

Sampling weights were used to account for the fact that the probabilities of selection were not identical for all students.

Scaling

Item response theory (IRT) methods were used to estimate average scale scores in CivEd for the nation as a whole and for various subgroups of interest. CivEd used two types of IRT models to estimate scale scores: the one-parameter Rasch model for the three civic achievement scales, and the Generalized Partial Credit model (GPC) for the attitudinal scales. The one-parameter Rasch model specifies the probability of a correct response as a logistic distribution in which items vary only in terms of their difficulty. This model is used on items that are scored correct or incorrect. The GPC model was developed for situation where item response are contained in two or more ordered categories (such as "agree" and "strongly agree"). Items are conceptualized as a series of ordered steps where examinees receive partial credit for successfully completing a step. The GPC is formulated based on the assumption that each probability of choosing the k^{th} category over the $(k-1)^{th}$ category is governed by the dichotomous (i.e., Rasch) response model.

Imputation

Imputation has not been performed.

Sampling Error

Because CivEd uses complex sampling procedures, it uses a Taylor series procedure to estimate standard errors.

Data Comparability

The CivEd International Coordinating Center (ICC), located at Humboldt University in Berlin, Germany, worked to ensure that the data collection procedures across countries are comparable. To this end, the ICC instituted the following procedures for quality assurance:

- Coordinated by the CivEd Sampling Referee, national school and student samples are rigorously reviewed for bias and international comparability.
- Utilizing two independent translations within each country, the CivEd materials are translated into the national languages of the participating countries. Once these translations are reconciled, the CivEd International Coordinating Center verifies these results through the use of a professional translation agency.
- Data collection staff from each nation are thoroughly trained in data collection and scoring procedures. Furthermore, the CivEd International Coordinating Center monitors the work of the national data collection staff throughout the entire project.
- Site visits by quality control staff are conducted during the testing period to further ensure the international data collection procedures are being followed at the national level.
- Finally, an extensive review of data is conducted for internal and cross-country consistency.

Within the United States, survey administrators discovered an unexpected problem in sampling classrooms within schools. They found that the increasing use of "block scheduling" in high schools created a situation where not all students within grade 9 were taking a given subject at the same time. Thus, while schools were able to provide a list of first-semester civics classes, not all students take civics during the first semester, even where civics is compulsory (some students can take civics during the second semester). Schools were also reluctant to assess students who had not yet taken civics, particularly if they were scheduled to take civics during the second semester, and schools also resisted drawing a sample of students from across more than one class. (The study had been promoted as assessing one classroom per school.)

Contact Information

For content information on the IEA Civics Study, contact:

Laurence Ogle
Phone: (202) 502–7426
E-mail: laurence.ogle@ed.gov

Mailing Address:

National Center for Education Statistics
1990 K Street NW
Washington, DC 20006–5651

Methodology and Evaluation Reports

Methodology discussed in technical notes.

Civic Education Study 1999 CD-ROM (NCES 2002–201). Washington, DC: 2002. [Includes the *1999 IEA Civic Education Study United States User's Guide* (NCES 2002–003), by Trevor Williams, Stephen Roey, Connie Smith, Dward Moore, David Kastberg, and Jean Fowler.]

What Democracy Means to Ninth-Graders: U.S. Results from the International IEA Civic Education Study, NCES 2001–096, by Stephane Baldi, Marianne Perie, Dan Skidmore, Elizabeth Greenberg, and Carole Hahn. Torney-Purta, John Schwille, and Jo-Ann Amadeo. Amsterdam, The Netherlands: 1999.

Civic Education Across Countries: Twenty-four National Case Studies from the IEA Civic Education Project, by Judith Torney-Purta, John Schwille, and Jo-Ann Amadeo. Amseterdam, The Netherlands: 1999.

Appendix A: Glossary of Statistical Terms

Balanced Incomplete Block (BIB) spiraling: In a BIB design, as in standard matrix sampling, no sample unit is administered all of the tasks in the assessment pool. However, unlike standard matrix sampling (in which items or tasks are assembled into discrete booklets), BIB design requires that sample units receive different interlocking sections of the assessment forms that allow for the estimation of relationships among all the tasks in the pool through the unique linking of blocks.

"Spiraling" refers to the method by which test booklets are assigned to sample units. Each version of the assessment booklet must appear in the sample approximately the same number of times and must be administered to equivalent subgroups within the full sample. To ensure proper distribution at assessment time, the booklets are packed in spiral order (e.g., one each of booklets 1 through 7, then 1 through 7 again, and so on). The test coordinator randomly assigns these booklets to the sample units in each test administration session. Spiraled distribution of the booklets promotes comparable sample sizes for each version of the booklet, ensures that these samples are randomly equivalent, and reduces the likelihood that sample units will be seated within viewing distance of an identical booklet.

Balanced Repeated Replication (BRR): See Replication techniques.

Base weight: The product of the reciprocals of the probabilities of inclusion for all stages of sampling.

Bias (due to nonresponse): The difference that occurs when respondents differ as a group from nonrespondents on a characteristic being studied.

Bias (of an estimate): The difference between the expected value of a sample estimate and the corresponding true value for the population.

Blanking edit: See Edits.

Bootstrap: See Replication techniques.

CAPI: Computer Assisted Personal Interviewing enables data collection staff to use portable microcomputers to administer a data collection form while viewing the form on the computer screen. As responses are entered directly into the computer, they are used to guide the interview and are automatically checked for specified range, format, and consistency edits.

CATI: Computer Assisted Telephone Interviewing uses a computer system that allows a telephone interviewer to administer a data collection form over the phone while viewing the form on a computer screen. As the interviewer enters responses directly into the computer, the responses are used to guide the interview and are automatically checked for specified range, format, and consistency edits.

Chi-squared Automatic Interaction Detector (CHAID) Analysis: This technique divides the respondent data into segments which differ with respect to the item being imputed. This segmentation process first divides the data into groups based on categories of the most significant predictors. It then splits each of these groups into smaller groups based on other predictor variables and merges categories of a variable found insignificant (by x^2 test). This splitting and merging progress continues until no more statistically significant predictors are found. The imputation classes form the final CHAID segments.

Cohort: A group of individuals who have a statistical factor in common (e.g., year of birth, grade in school, year of high school graduation).

"Cold-deck" imputation: See Imputation.

Component weight: For each stage of sampling, the component weight is equal to the reciprocal of the probability of selecting the unit at that stage.

Composite variables: Variables constructed through the combination of two or more variables (e.g., socioeconomic status) or through calculation by applying a mathematical function to a variable. Composite variables are also referred to as *derived, constructed,* or *classification* variables.

Computer Assisted Personal Interviewing: See CAPI.

Computer Assisted Telephone Interviewing: See CATI.

Consistency edits: See Edits.

Coverage error: Coverage error in an estimate results from the omission of part of the target population (undercoverage) or the inclusion of units from outside the target population (overcoverage).

Critical items or key items: Items deemed crucial to the methodological or analytical objectives of the study.

Dependent variable: A mathematical variable whose value is determined by that of one or more other variables in a function. In regression analysis, when a random variable, y, is expressed as a function of variables X_1, X_2,..., plus a stochastic term, then y is known as the "dependent variable."

Design effect: The cumulative effect of the various design factors affecting the precision of statistics is often modeled as the sample design effect. The design effect, DEFF, is defined as the ratio of the sampling variance of the statistic (e.g., a mean or a proportion) under the actual sampling design divided by the variance that would be expected for a simple random sample of the same size. Hence, the design effect is equal to one, by definition, for simple random samples. For clustered multistage sampling designs, the design effect is greater than unity, reflecting that the precision is less than could be achieved with a simple random sample of the same size (if that were the sampling design). The size of the design effect depends largely on the intracluster correlation of the survey observations within the primary sampling units. Hence, statistics that are based on observations that are highly correlated within units will have higher design effects.

Durbin's Method: This method selects two first-stage units per stratum without replacement, with probability proportional to size so that the joint inclusion probabilities are greater than zero for every pair.

Edits: These are procedures for checking and modifying response in a survey.

> ***Blanking edit*:** Deletes extraneous entries and assigns the "not answered" code to items that should have been answered but were not.
>
> ***Consistency edit*:** Identifies inconsistent entries within each record and, whenever possible, corrects them. If they cannot be corrected, the entries are deleted. Inconsistencies can be (1) within items or (2) between items. The consistency edit also fills some items where data are missing or incomplete by using other information on the data record.
>
> ***Logic edit*:** Checks made of the data to ensure logical consistency among the responses from a data provider.
>
> ***Range check*:** Determines whether responses fall within a predetermined set of acceptable values.
>
> ***Relational edit check*:** Compares data entries from one section of the questionnaire for consistency with data entries from another section of the questionnaire.
>
> ***Skip pattern check*:** Checks if responses correctly followed skip pattern instructions.
>
> ***Structural edit check*:** Checks that each case has the correct segments.

Summation check: Compares reported totals with the sums of the constituent data items.

Estimate: A numerical value obtained from a statistical sample and assigned to a population parameter. The particular value yielded by an estimator in a given set of circumstances or the rule by which such particular values are calculated.

Estimation: Estimation is concerned with inference about the numerical value of unknown population values from incomplete data, such as a sample. If a single figure is calculated for each unknown parameter, the process is called point estimation. If an interval is calculated within which the parameter is likely, in some sense, to lie, the process is called interval estimation.

Field test: The study of a data collection activity in the setting where it is to be conducted.

"Hot-deck" imputation: See Imputation.

Imputation: Imputation (for item or survey nonresponse) involves supplying a value if an item response is missing. The items may be missing because the respondent was careless, refused to provide an answer, or could not obtain the requested information. Since extensive amounts of missing data can seriously bias sample-based estimates, procedures for imputing missing values are often developed. Imputation is used to reduce nonresponse bias in survey estimates, simplify analyses, and improve the consistency of results across analyses.

Depending on the type of data to be imputed and the extent of missing values, a number of alternative techniques can be employed. These techniques include: logical imputation, the use of poststratum averages, "hot deck" imputation, and regression and other "modeling" techniques.

"Cold-deck" imputation: A process that imputes missing data with values observed from a past survey.

"Hot-deck" imputation: Hot deck refers to a general class of procedures for which cases with missing items are assigned the corresponding value of a "similar" respondent in the sample.

Random within class: This method divides the total sample into imputation classes according to the values of the auxiliary variables. Each nonrespondent is assigned a value randomly selected from the same imputation class.

Sequential (also known as traditional): The records of the survey are treated sequentially in the same imputation class and for each class a single value is stored to provide a starting point for a single pass through the data file. If a record has a response, that value replaces the previous value. If the record is missing, the currently stored value is assigned to that unit.

Logical imputation: Logical imputation can be applied in situations where a missing response can be inferred with certainty (or high degree of probability) from other information in the data record.

Poststratum averages: In the use of poststratum averages, a record with missing data is assigned the mean value of those cases in the same "poststratum" for which information on the item is available.

Proc Impute: This is an advanced software package that performs three steps for each target variable to be imputed:

1) Uses stepwise regression analysis to find the best combination of predictors among all variables included in the imputation model;
2) Creates homogeneous cells of records which have close predicted regression values; and
3) Imputes each missing record in a given cell with a weighted average of two donors, one from its own cell and the other from an adjacent cell.

Regression and other modeling techniques: These techniques operate by modeling the variable to be imputed, Y as a function of related independent variables, $X_1, X_2, \ldots, X_p$. To preserve the variability of the Ys at specific values of $X_1, \ldots, X_p$, a residual, $\hat{e}$, is sometimes added to the predicted value determined from the model.

Independent variable: In regression analysis, when a random variable, y, is expressed as a function of variables X_1, X_2,..., plus a stochastic term, the Xs are known as "independent variables."

Item nonresponse: An item on a data collection form that is missing when a response was expected.

Jackknife method: See Replication techniques.

Key items or critical items: Items deemed crucial to the methodological or analytical objectives of the study.

Keyfitz approach: A method of probability selection that maximizes the selected units from a past sample.

Logic edit: See Edits.

Logical imputation: See Imputation.

Measurement error: Measurement error refers to errors in estimates resulting from incorrect responses gathered during the data collection phase of a survey. Measurement errors result, for instance, when the respondent gives (intentionally or unintentionally) incorrect answers, the interviewer misunderstands or records answers incorrectly, the interviewer influences the responses, the questionnaire is misinterpreted, etc.

Mitofsky-Waksberg method: A method of sample selection for household telephone interviewing via random digit dialing where the sampling is carried out through a two-stage design. As Waksberg explained in his 1978 *Journal of the American Statistical Association* article: "Obtain from AT&T a recent list of all telephone area codes and existing prefix numbers within the areas. To these add all possible choices for the next two digits, and thus prepare a list of all possible first eight digits of the ten digits in telephone numbers. These eight-digit numbers are treated as Primary Sampling Units (PSU). A random selection is then made of an eight-digit number, and also of the next two digits. The number is then dialed. If the dialed number is at a residential address, the PSU is retained in the sample. Additional last two digits are selected at random and dialed within the same eight-digit group, until a set number, k, of residential telephones is reached. Interviews are attempted both at the initial number and the additional k numbers. If the original number called was not residential, the PSU is rejected. This process is repeated until a predesignated number of PSU's, m, is chosen. The total sample size is, therefore, $m(k + 1)$. The values of m and k are chosen to satisfy criteria for an optimum sample design." Note that although all units have the same probabilities of selection, it is not necessary to know the probabilities of selection of the first-stage or the second-stage units.

Nonresponse: Cases in data collection activities in which potential data providers are contacted but refuse to reply or are unable to do so for reasons such as deafness or illness.

Nonresponse bias: This occurs when respondents as a group differ from nonrespondents in their answers to questions on a data collection form.

Nonsampling error: This term is used to describe variations in the estimates that may be caused by population coverage limitations, as well as data collection, processing, and reporting procedures. The sources of nonsampling errors are typically problems like unit and item nonresponse, the differences in respondents' interpretations of the meaning of the questions, response differences related to the particular time the survey was conducted, and mistakes in data preparation.

Open-ended: A type of interview question that does not limit the potential response to predetermined alternatives.

Ordinary least squares (OLS): The estimator that minimizes the sum of squared residuals.

Out-of-range response: A response that is outside of the predetermined range of values considered acceptable for a particular item.

BEST COPY AVAILABLE

Oversampling: Deliberately sampling a portion of the population at a higher rate than the remainder of the population.

Plausible values: Proficiency values drawn at random from a conditional distribution of a survey respondent, given his or her response to cognitive exercises and a specified subset of background variables.

Plausible values methodology: Plausible values methodology represents what the true performance of an individual might have been, had it been observed, using a small number of random draws from an empirically derived distribution of score values based on the student's observed responses to assessment items and on background variables. Each random draw from the distribution is considered a representative value from the distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. The draws from the distribution are different from one another to quantify the degree of precision (the width of the spread) in the underlying distribution of possible scale scores that could have caused the observed performances.

Population: All individuals in the group to which conclusions from a data collection activity are to be applied.

Population variance: A measure of dispersion defined as the average of the squared deviations between the observed values of the elements of a population and the corresponding mean of those values.

Poststratification: An estimation method that adjusts the sampling weights so that they sum to specified population totals corresponding to the levels of a particular response variable.

Poststratification adjustment: A weight adjustment that forces survey estimates to match independent population totals within selected poststrata (adjustment cells).

Precision: The difference between a sample-based estimate and its expected value. Precision is measured by the sampling error (or standard error) of an estimate.

Pretest: A test to determine performance prior to the administration of a data collection activity.

Probability sample: A sample selected by a method such that each unit has a fixed and determined probability of selection.

Proc Impute: See Imputation.

Processing: The manipulation of data.

Range check: See Edits.

Regression analysis: A statistical technique for investigating and modeling the relationship between variables.

Relational edit check: See Edits.

Replicate estimate: An estimate of the population quantity based on the replicate subsample using the same estimation methods used to compute the full sample estimate.

Replicate sample: One of a set of subsamples, each obtained by deleting a number of observations in the original sample for the purpose of computing the appropriate variance based on the complex design of the survey.

Replicate weight: The weight assigned to an observation for a particular replicate subsample.

Replicates: A term often used to refer to either the replicate sample or the replicate estimate, depending on context.

Replication techniques: Methods of estimating sampling errors that involve repeated estimation of the same statistic using various subsets of data providers. The major methods are balanced repeated replication (BRR), bootstrap, and the jackknife technique.

BEST COPY AVAILABLE

Balanced Repeated Replication (BRR): A method of replication that divides the sample into half-samples.

Bootstrap: A resampling technique of creating replicates by drawing random samples with replacement that mirror the original sampling plan for a pseudo-population constructed from the original sample.

Jackknife method: A method of replication that creates replicates (subsets) by excluding one unit at a time from the sample.

Sample: A subgroup selected from the entire population.

Sampling error: When a sample rather than the entire population is surveyed, estimates can differ from the true population values that they represent. This difference, or sampling error, occurs by chance, and its variability is measured by the standard error of the estimate. Sample estimates from a given survey design are unbiased when an average of the estimates from all possible samples would yield, hypothetically, the true population value. In this case, the sample estimate and its standard error can be used to construct approximate confidence intervals, or ranges of values, that include the true population value with known probabilities.

Sampling variance: A measure of dispersion of values of a statistic that would occur if the survey were repeated a large number of times using the same sample design, instrument, and data collection methodology. The square root of the sampling variance is the standard error.

Sampling weights: See Weighted estimates.

Scaling (item response theory): Item response theory (IRT) scaling assumes some uniformity in response patterns when items require similar skills. Such uniformity can be used to characterize both examinees and items in terms of a common scale attached to the skills, even when all examinees do not take identical sets of items. Comparisons of items and examinees can then be made in reference to a scale, rather than to the percent correct. IRT scaling also allows the distributions of examinee groups to be compared.

This is accomplished by modeling the probability of answering a question in a certain ways as a mathematical function of proficiency or skill. The underlying principle of IRT is that, when a number of items require similar skills, the regularities observed across patterns of response can often be used to characterize both respondents and tasks in terms of a relatively small number of variables. When aggregated through appropriate mathematical formulas, these variables capture the dominant features of the data. IRT enables the assessment of a sample of students in a subject area or subarea on a common scale even if different students have been administered different exercises.

Skip pattern check: See Edits.

Special population: A subset of the total population distinguishable by unique needs, characteristics, or interests (e.g., disadvantaged students, gifted students, physically or mentally handicapped students, vocational education students).

Spiraling: See Balanced Incomplete Block (BIB) spiraling.

Standard deviation: The most widely used measure of dispersion of a set of values. It is equal to the positive square root of the population variance.

Standard error: The positive square root of the sampling variance. It is a measure of the dispersion of the sampling distribution of a statistic. Standard errors are used to establish confidence intervals for the statistics being analyzed.

Statistically significant: There is a low probability that the result is attributable to chance alone.

Structural edit checks: See Edits.

Summation check: See Edits.

Taylor Series: The Taylor Series variance estimation procedure is a technique to estimate the variances of nonlinear statistics. The procedure takes the first-order Taylor Series approximation of the nonlinear statistic and then substitutes the linear representation into the appropriate variance formula based on the sample design. For stratified multistage surveys such as the National Postsecondary Student Aid Study (NPSAS), the Taylor Series procedure requires analysis strata and analysis replicates defined from the sampling strata and primary sampling units (PSUs) used in the first stage of sampling.

Target population: See Population.

Time series: A set of ordered observations on a quantitative characteristic of an individual or collective phenomenon taken at different points in time. Usually the observations are successive and equally spaced in time.

Unit nonresponse: The failure of a survey respondent to provide any information.

Variable: A quantity that may assume any one of a set of values.

Variance: See Population variance and Sampling variance.

Weighted estimates: Estimates from a sample survey in which the sample data are weighted (multiplied) by factors reflecting the sample design. The weights (referred to as sampling weights) are typically equal to the reciprocals of the overall selection probabilities, multiplied by a nonresponse or poststratification adjustment.

Appendix B: Ordering NCES Publications and Data Files

Much NCES data and many NCES publications are available through the NCES web site. The NCES Electronic Catalog (http://nces.ed.gov/pubsearch/) allows searching for NCES products by NCES number or, for products released within the last 5 years, by keyword, survey/program area, type of product, and release date. The Electronic Catalog also has lists of publications published in the last 90 days, data products released in the last 6 months, and lists of all publications and data products by survey and program area.

In addition to downloading from the NCES web site, there are three other ways to obtain NCES publications, CD-ROMs, and other products:

1. Education Publications Center (ED Pubs),
2. Government Printing Office (GPO), and
3. Federal Depository Libraries.

Until supplies are exhausted, a single copy of a publication or CD-ROM may be obtained at no cost from ED Pubs. Before requesting a copy, it is necessary to have the complete title and NCES number for the publication; for example, *The Condition of Education, 2002*, NCES 2002–025.

Education Publications Center (ED Pubs)
Toll-free number: (877) 4ED–Pubs, (877) 433–7827
TTY/TDD toll-free number: (877) 576–7734
Fax: (301) 470–1244
E-mail: customerservice@edpubs.org
Internet: www.edpubs.org/

Mailing Address:
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

If more than one copy of a publication is needed, or if ED Pubs' supplies have been exhausted, many—not all—NCES products may be purchased from the Government Printing Office (GPO). To order a copy from GPO, it is necessary to have the product's GPO stock number (e.g., 065-000-00871-8). The product's stock number and price can be found out by going to the U.S. Government Online Bookstore and entering the product's title or key words.

Government Printing Office (GPO)
Online orders: http://bookstore.gpo.gov/
Phone orders: 1–866–512–1800 (toll-free); (202) 512–1800 (DC area)
FAX: Credit card orders may be faxed to (202) 512–2250

Mailing Address:
Superintendent of Documents
P.O. Box 371954
Pittsburgh, PA 15250–7954

Federal Depository Library

For older publications, the only source for an NCES publication may be a Federal depository library. There are approximately 1,350 of these libraries around the country. However, only the "Regional" libraries receive all materials distributed through the Federal Depository Library Program. Other Federal depository libraries select materials according to the needs of their communities. These libraries can be located through the following web site:

http://www.gpo.gov/su_docs/locators/findlibs/index.html

Public-use versus Restricted-use Data Files

NCES uses the term "public-use data" for survey data when the individually identifiable information has been coded or deleted to protect the confidentiality of survey respondents. All NCES public-use data files can be accessed (at no cost) from the NCES web site. Only public-use data files that are on CD-ROM are available through ED Pubs or GPO.

Restricted-use data files contain individually identifiable information, which is confidential and protected by law. To use these data, researchers must obtain a restricted-use data license. A brief summary of the steps that need to be followed to obtain (or amend) a restricted-use data license is provided below. The procedures are fully discussed in the *NCES Restricted-Use Data Procedures Manual* (NCES 96-860, http://nces.ed.gov/statprog/rudman/). To obtain a restricted data license (or to amend an existing license), the researcher must request the data in a letter addressed to the NCES Data Security Office.

Mailing Address:

Data Security Office
Department of Education/NCES/ODC
1990 K Street NW
Washington, DC 20006

The letter will need to include the following:

1. The license number to be amended (if the researcher already has a license);
2. The name of the dataset(s) the researcher wishes to use;
3. The purpose for the loan of the data;
4. The length of time the researcher will need the data;
5. The computer security plan the researcher will follow;
6. The list of authorized users; and
7. An affidavit of nondisclosure for each authorized user, promising to keep the data completely confidential.

A researcher who is amending an existing license and whose purpose is a continuation of the project that was approved originally may be able to condense the abstract of the research design, but the description must be specific enough to justify using the raw data. Similarly, researchers who plan to use the same computer(s) as person(s) who are already licensed users may be able to simply update the computer security plan previously approved. Computer security plans need to be followed carefully as spot site inspections do occur. In the case of postsecondary institutions, only faculty can serve as the primary project officer. Graduate students may be listed as authorized users only.

Contact Person:

Cynthia L. Barton
Data Security Assistant
Phone: (202) 502–7307
E-mail: cynthia.barton@ed.gov

Note on Working Papers: Working papers are available on the NCES web site through the Electronic Catalog.

Appendix C: Web-based and Standalone Tools for Use with NCES Survey Data

NCES has developed a number of web-based and standalone tools for use with its data.* There are two user tools that have been developed for use across multiple surveys: data analysis systems (DAS), which produce tabular data for the user; and electronic codebooks (ECBs), which allow users to develop datafiles in SAS, SPSS, or ASCII format. These are described in more detail below, along with a list of the surveys available with each. Following this, descriptions of the tools developed for more specialized uses—for example, the Private School Locator and the NAEP Test Questions Tool—are provided in a survey-by-survey list.

DAS (http://nces.ed.gov/das/): The Data Analysis System (DAS) is a software application that provides access to Department of Education survey data. DAS allows users to create programming instruction files (DAS files) that specify the information they want displayed in a table. The output table will contain the estimates (usually percentages of students) and corresponding standard errors which are calculated taking into account the complex sampling designs used in NCES surveys. In addition, the DAS software can create correlation matrices which can be used as input for most popular statistical programs to conduct multivariate analysis. There is a separate DAS for each survey data set, and all have a consistent interface and command structure. DAS applications are available in Windows- and web-based formats. The available surveys are:

- Baccalaureate and Beyond (B&B) Longitudinal Study
- Beginning Postsecondary Students (BPS) Longitudinal Study
- High School and Beyond (HS&B) Longitudinal Study
- National Education Longitudinal Study of 1988 (NELS:88)
- National Household Education Surveys (NHES) Program
- National Longitudinal Study of the High School Class of 1972 (NLS-72)
- National Postsecondary Student Aid Study (NPSAS)
- National Study of Postsecondary Faculty (NSOPF)

Electronic Codebook (ECB) programs have been created for many NCES surveys. These programs, after being installed on a user's personal computer, allow the user to examine the variables in each of a survey's data files, as well as create SAS and SPSS programs that will generate an extract data file from any of the survey data files on the CD-ROM.

ECB programs are usually included on a CD-ROM with the survey data, but NCES has issued a CD-ROM that contains only electronic codebooks. This CD was created to provide updated ECB software for data sets that were, in some cases, released several years ago.

ECBs may be available for use with public-use data, restricted-use data, or both, depending on the survey. ECBs are available for the following surveys:

- Baccalaureate and Beyond (B&B) Longitudinal Study—Restricted-use

---

*As explained in appendix B, all NCES public-use data files can be accessed (at no cost) from the NCES web site. To use restricted-use data, researchers must first obtain a restricted-use data license.

BEST COPY AVAILABLE

- Beginning Postsecondary Students (BPS) Longitudinal Survey—Restricted-use
- Early Childhood Longitudinal Study (ECLS)—Public-use and Restricted-use
- High School and Beyond (HS&B) Longitudinal Study—Restricted-use
- High School Transcript Study—Restricted-use
- Integrated Postsecondary Education Data System (IPEDS)—Public-use
- National Adult Literacy Survey (NALS)—Restricted-use
- National Education Longitudinal Study of 1988 (NELS:88)—Public-use and Restricted-use
- National Household Education Surveys (NHES) Program—Public-use and Restricted-use
- National Longitudinal Study of the High School Class of 1972 (NLS-72)—Public-use
- National Postsecondary Student Aid Study (NPSAS)—Restricted-use
- National Study of Postsecondary Faculty (NSOPF)—Public-use and Restricted-use
- Private School Universe Survey (PSS)—Public-use
- Schools and Staffing Survey (SASS)—Public-use and Restricted-use

EARLY CHILDHOOD EDUCATION SURVEY

Early Childhood Longitudinal Study (ECLS)

- ECB for ECLS—Public-use and Restricted-use

ELEMENTARY AND SECONDARY EDUCATION SURVEYS

Common Core of Data (CCD)

- **CCD CD-ROM Interface:** After selecting one of the three databases—School, Agency, or State—the user enters search criteria in specific fields, in order to limit the number of records for review to a select group. These records (matching the search criteria) can be displayed in summary or detail format, and can be printed. Specific fields for the selected records may be chosen and data exported to be used with other software packages. There are a number of export formats available that can be used with spreadsheets, databases, word processing packages, and statistical software packages.

- **National Public School and School District Locator** (http://nces.ed.gov/ccd/search.asp): The School/District Locator enables users to find the correct name, address, telephone number, NCES ID number, locale (rural, large city, etc.), and other student and teacher information for public schools or school districts for the latest school year as reported to NCES by state education officials in each state. The Locator includes a Locator Glossary, which includes variable codes and definition descriptions, and a list of newly reported schools and school districts (this information is from unedited state data submissions and is updated as new information is received).

- **Public School District Finance Peer Search** (http://nces.ed.gov/edfin/search/search_intro.asp): This search allow users to compare the finances of a school district with its peers (those districts which share similar characteristics to the one chosen). Users may enter the entire name or only a portion of it. If more than one district with that name is found, users are prompted to select the appropriate one. Once the user has narrowed the search to one district, peer districts will be selected based on: enrollment, student/teacher ratio, median household income, district type, and metro status location. Users can base their search for peers on a different set of criteria using the "Advanced" feature. Users wishing to perform a search other than a peer search may use the "Expert" feature.

BEST COPY AVAILABLE

Private School Universe Survey (PSS)

- **ECB for PSS**—Public-use

- **Private School Locator** (http://nces.ed.gov/surveys/pss/privateschoolsearch/): The Private School Locator enables users to find school names, addresses, and other school information for private schools. Information for a particular private school, or a group of private schools, can be retrieved based on selection criteria the user specifies. Users can also download the entire Locator data base (2.3 MB), or download an ASCII text data file of the schools selected once the selection process is completed. The information in this locator comes from the approximately 29,000 schools that participated in the latest NCES Private School Survey (PSS). Users can request that schools not found in the Locator be included in future PSS. The Locator is also available on CD-ROM.

Schools and Staffing Survey (SASS)

- **ECB for SASS**—Public-use and Restricted-use

- **Schools and Staffing Survey (SASS) Item Bank** (http://nces.ed.gov/surveys/sass/sassib/): The Item Bank provides the opportunity to search and view all items that appear in the 1993–94 and 1999–2000 SASS questionnaires and the 1994–95 Teacher Follow-up Survey (TFS) questionnaires. The Content Framework is an outline of topics surveyed by SASS and TFS. In addition to searching for items, the Item Bank allows users to print detailed lists of items from the questionnaires; for example, the results of a search on a particular keyword.

National Education Longitudinal Study of 1988 (NELS:88)

- **DAS for NELS:88**—See DAS, above.

- **ECB for NELS:88**—Public-use and Restricted-use

National Longitudinal Study of the High School Class of 1972 (NLS-72)

- **DAS for NLS-72**—See DAS, above.

- **ECB for NLS-72**—Public-use

High School and Beyond (HS&B) Longitudinal Study

- **DAS for HS&B**—See DAS, above.

- **ECB for HS&B**—Restricted-use

LIBRARY SURVEYS

Public Libraries Survey (PLS)

- **Public Library Locator** (http://nces.ed.gov/surveys/libraries/liblocator/): This tool helps users locate information about a public library or a public library service outlet when users know some, but not all of the information about it. The information in this locator has been drawn from the NCES Public Libraries Survey.

- **Public Library Peer Comparison Tool** (http://nces.ed.gov/surveys/libraries/publicpeer/): This tool allows the user to get information on a particular library, or to customize a peer group by selecting the key variables that are used to define it. The user can then view customized reports of the comparison between the library of interest and its peers, on a variety of variables as selected by the user. There is a tutorial for this tool.

Academic Library Survey (ALS)

- **Academic Library Peer Comparison Tool** (http://nces.ed.gov/surveys/libraries/academicpeer/): This tool allows the user to get information on a particular library, or to customize a peer group by selecting the key variables that are used to define it. The user can then view customized reports of the comparison between the library of interest and its peers, on a variety of variables as selected by the user.

BEST COPY AVAILABLE

POSTSECONDARY AND ADULT EDUCATION SURVEYS

Integrated Postsecondary Education Data System (IPEDS)

- ECB for IPEDS—Public-use
- IPEDS College Opportunities On-line—COOL (http://nces.ed.gov/ipeds/cool/): This is a direct link to over 9,000 colleges and universities in the United States. It was developed after NCES was authorized by Congress in 1998 to help college students, future students, and their parents understand the differences between colleges and how much it costs to attend college. Users can name a specific college or set of colleges and obtain information about them or use the search feature to find a college based on its location, program, or degree offerings either alone or in combination. The more criteria the user specifies, the smaller the number of colleges that will fit the criteria. Once the user has identified some colleges of interest, he or she can obtain important and understandable information on all of them.
- IPEDS Peer Analysis System (Postsecondary Institutions) (http://nces.ed.gov/ipedspas/): This tool is designed to enable a user to easily compare a postsecondary institution of the user's choice to a group of peer institutions, also selected by the user. This is done by generating reports using selected IPEDS variables of interest. There are tutorials for this tool.

National Study of Postsecondary Faculty (NSOPF)

- DAS for NSOPF—See DAS, above.
- ECB for NSOPF—Public-use and Restricted-use

National Postsecondary Student Aid Study (NPSAS)

- DAS for NPSAS—See DAS, above.
- ECB for NPSAS—Restricted-use

Beginning Postsecondary Students (BPS) Longitudinal Study

- DAS for BPS—See DAS, above.
- ECB for BPS—Restricted-use

Baccalaureate and Beyond (B&B) Longitudinal Study

- DAS for B&B—See DAS, above.
- ECB for B&B—Restricted-use

EDUCATIONAL ASSESSMENT SURVEYS

National Assessment of Educational Progress (NAEP)

- NAEP Data Tool Kit, including NAEPEX: This is a data extraction program for choosing variables, extracting data, and generating SAS and SPSS control statements, and analysis modules for cross-tabulation and regression that work with SPSS and Excel (available on CD-ROM).
- NAEP Test Questions Tool (http://nces.ed.gov/nationsreportcard/itmrls/): The purpose of this tool is to provide easy access to NAEP questions, actual student responses, and scoring guides used in released portions of the NAEP assessments. National and, where appropriate, state data are also presented. Note that entire NAEP assessments are not presented here, since some questions must be kept secure for use in future NAEP assessments. Science is currently available only as a PDF document. There is a tutorial for this tool.
- NAEP State Profiles (http://nces.ed.gov/nationsreportcard/states/): The State Profiles present key data about each state's student and school population and its NAEP testing history and results. The profiles also contain links to other sources of information on the NAEP web site, including the most recent state report cards for all available subjects.

BEST COPY AVAILABLE

- NAEP Data Tool (http://nces.ed.gov/nationsreportcard/naepdata/): This tool provides users with tables of detailed results from NAEP's national and state assessments. The data are based on information gathered from the students, teachers, and schools that participated in NAEP. There is a tutorial for this tool.

Third International Mathematics and Science Study (TIMSS)

- TIMSS Videotape Classroom Study CD-ROM: Actual footage of 8th-grade mathematics classes lets viewers see first hand an abbreviated geometry and algebra lesson in each of three countries: Germany, Japan, and the United States.

National Adult Literacy Survey (NALS)

- ECB for NALS—Restricted-use

HOUSEHOLD SURVEYS

National Household Education Surveys (NHES) Program

- DAS for NHES—See DAS, above.
- ECB for NHES—Public-use and Restricted-use

SMALL SPECIAL-PURPOSE NCES SURVEYS

High School Transcript Study

- ECB for 1998 High School Transcript (HST) Study—Restricted-use
- Table Generator (TabGen) software: A simplified version of WesVar software, TabGen computes estimates and replicate variance estimates for collected data and displays its results in Microsoft Excel workbooks. Users can create tables that display frequencies, percentages, means, medians, standard errors, quantiles, confidence intervals, coefficients of variance, and more.

Appendix D: NCES Survey Web Site Addresses

Every effort has been made to verify the accuracy of all URLs listed in this Handbook at the time of publication. If a URL is no longer working, try using the root directory to search for a page that may have moved. For example, if the link to http://nces.ed.gov/surveys/libraries/academic.asp is not working, try *http://nces.ed.gov/* and search the NCES Site Index for Academic Libraries.

Survey	**Web site**
Academic Libraries Survey (ALS)	http://nces.ed.gov/surveys/libraries/academic.asp
Baccalaureate and Beyond (B&B) Longitudinal Study	http://nces.ed.gov/surveys/b&b/
Beginning Postsecondary Students (BPS) Longitudinal Study	http://nces.ed.gov/surveys/bps/
Civic Education Study (CivEd)	http://nces.ed.gov/surveys/cived/
Common Core of Data (CCD)	http://nces.ed.gov/ccd/
Current Population Survey (CPS), October and September Supplements	http://nces.ed.gov/surveys/cps/
Early Childhood Longitudinal Study (ECLS)	http://nces.ed.gov/ecls/
Fast Response Survey System (FRSS)	http://nces.ed.gov/surveys/frss/
Federal Library Survey (FedLib)	http://nces.ed.gov/surveys/libraries/federal.asp
High School and Beyond (HS&B) Longitudinal Study	http://nces.ed.gov/surveys/hsb/
High School Transcript (HST) Studies	http://nces.ed.gov/surveys/hst/
IEA Reading Literacy Study	See http://nces.ed.gov/surveys/pirls/
Integrated Postsecondary Education Data System (IPEDS)	http://nces.ed.gov/ipeds/
International Adult Literacy Survey (IALS)	See http://nces.ed.gov/surveys/all/
Library Cooperatives Survey (LCS)	http://nces.ed.gov/surveys/libraries/coops.asp

Survey	**Web site**
National Assessment of Educational Progress (NAEP)	http://nces.ed.gov/nationsreportcard/
National Assessments of Adult Literacy (NAAL), including the 1992 National Adult Literacy Survey (NALS)	http://nces.ed.gov/naal/
National Education Longitudinal Study of 1988 (NELS:88)	http://nces.ed.gov/surveys/nels88/
National Household Education Surveys (NHES) Program	http://nces.ed.gov/nhes/
National Longitudinal Study of the High School Class of 1972 (NLS-72)	http://nces.ed.gov/surveys/nls72/
National Postsecondary Student Aid Study (NPSAS)	http://nces.ed.gov/surveys/npsas/
National Study of Postsecondary Faculty (NSOPF)	http://nces.ed.gov/surveys/nsopf/
Postsecondary Education Quick Information System (PEQIS)	http://nces.ed.gov/surveys/peqis/
Private School Universe Survey (PSS)	http://nces.ed.gov/surveys/pss/
Public Libraries Survey (PLS)	http://nces.ed.gov/surveys/libraries/public.asp
School Crime Supplement (SCS)	See http://www.ojp.usdoj.gov/bjs/
School Library Survey (SLS)	http://nces.ed.gov/surveys/libraries/school.asp
School Survey on Crime and Safety (SSOCS)	http://nces.ed.gov/surveys/ssocs/
Schools and Staffing Survey (SASS)	http://nces.ed.gov/surveys/sass/
State Library Agencies (StLA) Survey	http://nces.ed.gov/surveys/libraries/sla.asp
Survey of Earned Doctorates (SED)	http://www.nsf.gov/sbe/srs/ssed/
Teacher Follow-up Survey (TFS)	See http://nces.ed.gov/surveys/sass/
Third International Mathematics and Science Study (TIMSS)	http://nces.ed.gov/timss/

Appendix E: Index

The Index entries include survey component names and the words defined in the Key Concepts sections. Survey component names are italicized, and the page number where the component description appears in the Overview section is also italicized. Words defined in the Key Concepts sections are identified by an asterisk, and an asterisk follows the page number where the definition appears. Chapter page numbers are provided in the footer at the bottom of the page.

A

Academic Libraries (*Survey*) 2, 100, *105*–107, 110, *125*, 128, 131–132

Academic library* 2, 106*–107, 100, 112*

Accuracy 16, 32, 43, 60, 64, 78, 101, 143, 147, 182–183, 190, 195, 199, 215, 237, 240, 247–248, 280, 282, 284, 292

Achievement levels* 56, 191*, 201–202

Adult education 3, 70–71, 232–233, 244, 255–257, 259–264, 268–270, 272–273

Adult Education and Lifelong Learning 255, 257, 264

Adult literacy 3, 187, 203, 231–233, 235, 238–239, 241–246, 249–253, 268, 301

Allied operations* 111–112*

Analysis of variance—see Variance

Area frame 28–33, 29, 39, 48, 95, 201

Assessment design 7, 10, 18, 194, 204, 214, 235, 246

Assignment of session types to schools 192–193

Attainment* 56, 68, 83, 161, 162*, 163, 180, 232, 244, 271–273, 285

Attendance pattern* 150*

Automatic interaction detector (AID)—see Chi-squared automatic interaction detector

Auxiliary information 60

B

Background questionnaire 208, 230, 232, 235–236, 238–240, 244, 247, 250, 252

Background Questionnaire 244—see also *Student Background Questionnaire*

Balanced half-sample replication (BHR) 43

Balanced incomplete block (BIB) spiraling 195, 235

Balanced Repeated Replication (BRR), 43, 88

Base weight 14, 174–175, 197, 218–219, 237, 244, 249, 263, 265, 283–284, 292, 295–296

Base Year Data (from NPSAS) *161*, 163, 169–*170*

Base Year Ineligible (BYI) Study 53–55, 57–59, 61

Base Year Survey 8, 53, 55, 57–59, 61–62, *67*, 71–72, 75–77, 81–*82*, 83–84, 86, 89, 91

Basic CPS 271, *272*, 273–276

Before- and After-School Programs and Activities 256–257, 264, 268

Beginning students—see First-time beginning students (FTBs)

Benchmarking 207–208, 210, 249

BIA school* 38*, 44, 48, 96—see also Charter school, Private school, Public school

Bias 12, 15–17, 42, 44, 50, 60–64, 76, 79, 89–90, 145, 154–155, 165, 200, 216, 227, 233, 237–238, 249–250, 256, 260–264, 268–269, 276, 281, 297–298, 303

BIB spiraling—see Balanced incomplete block (BIB) spiraling

Bibliographic service center* 116*

Bibliographic utility* 116*

Birth Certificates 6, 9

Birth Cohort Study 5, 6, 18

Blanking edit 30, 41, 49, 95

Bootstrap 42–43

Branch institution* 126*

Branch library* 106*

Bureau of Indian Affairs (BIA) school—see BIA school

C

CACE—see Computer-assisted coding and editing

CADE—see Computer-assisted data entry

CAPI—see Computer-assisted personal interviewing

Care Provider and Preschool Teacher Interviews 7

Carnegie classifications 138, 180

Carnegie unit 65, 83, 295

Catholic (schools) 28, 84, 192, 201

CATI—see Computer-assisted telephone interviewing

Census—see Universe survey

Centralized processing center* 116*

Certainty 9, 29–30, 39, 42, 71, 85, 95, 143, 151–152, 154, 192, 200, 213, 225, 294

Charter school* 19, 36, 38*–42—see also BIA school, Private school, Public school

Chi-squared automatic interaction detector (CHAID) analysis 14, 155

Child assessments 5–6, 12–13

CIP 73, 123, 126, 129

CIP code* 126*

Circulation transaction* 106*

Civic Involvement 256

Civic involvement 256–257, 261, 268

Classification of Institutions 132

Classroom Teacher Questionnaire 6

Clerical edit 49, 95

Clerical imputation 41, 50, 96

Coding 13, 17, 58, 73, 78, 87, 140–141, 154, 156, 164, 166, 174, 176–177, 182, 196, 201, 217, 227, 236, 238, 240, 247–248, 250–251, 262, 265, 269, 282, 295, 299

Cognitive data 237–238, 247, 250, 252

Cognitive development 10–11

Cognitive items 217, 238, 250, 252

Cognitive test battery* 56*

Cognitive Tests 53–56, 58–59, 61–63, 65, 69, 82, *84*, 297

Cohort 5–13, 15, 18, 56, 59, 62–63, 69, 72, 81–86, 88–89, 91–92, 123, 127, 130, 152–153, 155, 161–166, 169–177, 211, 230, 292, 299

Cold-deck (imputation) 101, 130, 142

Collections* 112*

Combined school* 28*, 41, 47, 220, 279—see also Elementary school and secondary school

Comparisons (chapter survey listed first)
    CCD —with the Early Estimates Survey 25
        —within CCD 24
    HS&B —with NLS-72 91
        —within HS&B 91
    IEA —with NAEP reading assessments 230
        —with other countries 230
    IPEDS —with HEGIS 133
        —with SED 133
    NAEP —Main national vs. main state comparisons 202
        —Short-term trends 202
        —Long-term trends 202
        —Linking to Non-NAEP assessments (IAEP, TIMSS) 202–203
        —with NALS 203
        —with IEA Reading Literacy Study 203
    NALS —with the 1985 Young Adult Literacy Assessment 241
        —with the 1993 GED 241
    NELS:88 —across respondent groups 64
        —across survey waves 64

—with NLS-72 and HS&B 64
NHES —of methodology with other household surveys 267
—of topical data 267
NLS-72 —between NLS-72 student data and PETS data 78
—with HS&B and NELS:88 78
NPSAS —with IPEDS data 158
NSOPF —with other surveys (IPEDS mostly) 146
PSS —with National Catholic Educational Association data 33
—with CPS 33
—within PSS 33
SCS —between 1989 and 1995 and later NCVS victimization items 289
—between 1995 and 1999 NCVS and SCS items 290
—with other related surveys 290
SED —with IPEDS 186

Completions (C) 123

Component 5, 9–10, 14, 17, 19–20, 29–32, 35–37, 41–42, 55, 57–58, 65–66, 70, 89–90, 93–96, 99, 105, 111, 115, 122–123, 128, 153, 156, 170, 172, 174, 179, 187–190, 195, 197, 199, 209, 223, 229, 231, 255, 270, 290, 293–296, 300

Composite variable 8, 56, 163

Computer-assisted coding and editing (CACE) 295

Computer-assisted data entry (CADE) 73–74, 87, 140, 153–157, 182, 295

Computer-assisted personal interviewing (CAPI) 12–13, 15, 164

Computer-assisted telephone interviewing (CATI) 12–13, 30, 40–41, 59–60, 86–87, 90–91, 96, 139–140, 143–144, 152–158, 163–164, 172–174, 261–262, 288

Confidence interval 33, 275

Consistency edit 13, 30, 41, 49, 95, 107, 182

Consolidated Form (CN and CN-F) 125

Content changes 24, 42, 78, 146, 230

Cooperative collection resource facility* 116*

Correlation coefficient 196

Counselor Questionnaire 68

Course offering and course taking* 84*

Course Offerings Component 55

Coverage error 2, 16, 23, 31, 44, 61, 75–76, 89, 102, 110, 114, 118, 131, 143, 156, 165, 175, 183, 201, 220, 229, 239, 251, 258, 265, 276, 281, 284, 289, 297

Critical item 59, 69, 72–73, 75, 77, 86–87, 140, 154, 182, 184

Cross-sectional survey 291

Curriculum Studies 209

ALS	105–110	Fed. Lib.	115–119	LCS	300–301	NSOPF	135–147	SLS	93–97
B&B	169–177	FRSS	279–283	NAEP	187–205	PEQIS	283–286	SSOCS	291–292
BPS	161–167	HS&B	81–92	NALS	231–242	PLS	99–104	StLA	111–114
CCD	19–25	HST	292–300	NELS:88	53–66	PSS	27–34	TFS	47–51
CivEd	301–304	IALS	243–253	NHES	255–270	SASS	35–46	TIMSS	207–222
CPS	271–277	IEA	223–230	NLS-72	67–79	SCS	287–291		
ECLS	5–18	IPEDS	121–134	NPSAS	149–159	SED	179–186		

D

Data collection and processing 1–2, 11, 21, 25, 29, 40, 49, 58, 61, 72, 75, 85, 95, 101, 107, 112, 117, 127, 139, 143, 153, 163, 172, 181, 195, 211, 215, 226, 235, 247, 261, 274, 280, 284, 288, 291, 294, 301–302

Data comparability 17, 24, 32, 51, 64, 77, 91, 102, 104, 114, 118, 131–133, 146, 158, 177, 185, 202, 221, 230, 241, 252, 267, 282, 285, 289, 298, 303

Data entry 41, 59–60, 73, 77, 79, 87, 101, 113, 129, 140–141, 150, 153–154, 158, 182, 196, 217, 227, 265, 282, 295

Data processing 1, 21, 59, 140, 163, 173, 182, 195, 217, 221, 226, 236, 247, 251, 265, 275

Definitional differences 21, 23, 103, 146

Degree-granting institution* 126*, 171*

Department Chairperson Survey 136, 140, 144, 146

Department of Defense (schools) 19, 21, 38, 42, 117, 192, 201, 302

Department of Education Administrative Records 150

Dependency level* 171*

Dependency status* 150*—see also Dependency level

Design changes 41, 43, 77, 146, 158, 230

Design effects 15, 17, 43–44, 46, 61, 75, 88, 152, 165, 229, 298

Development of framework and questions 194

Direct Child Assessments 5, 12

Doctorate-granting institution* 180*

Document literacy* 188, 233*, 241, 244*, 246

Documents 59, 73, 110, 112, 163, 196–197, 223–224, 232–233, 236, 240, 246, 295

Documents* (i.e., "types of text") 224*

Double sampling 193

Dropout* 20, 23–24, 54, 56*, 58–61, 66, 68, 82–83, 89, 189, 266, 269–270, 272, 299

Durbin's method 9

E

Early Childhood Education and School Readiness 2, *256–257*, 259–261, 263, 266–267

Early Estimates Survey 19–*20*, 22–23, 25, 27

Edit check 22, 60, 88, 101, 107-108, 110, 113, 117, 154, 174

Editing 13, 22, 25, 30, 49, 58–60, 73, 77, 87, 95, 107, 113, 117, 127, 129–130, 132, 140–141, 154–155, 164, 174, 176, 181–182, 196, 217, 227, 236, 248, 262, 265, 269, 274, 280, 282, 284, 295—see also Processing

Educational attainment* 56, 68, 83, 163, 180, 232, 244, 271–273*

Effective sample size 212, 218

Elementary and Secondary School Students Survey 187–*188*, 191

Elementary school* 10, 12, 28*, 53, 55, 260, 271–273, 279, 281–282, 291—see also Combined school and Secondary school

ALS	105–110	Fed. Lib	115–119	LCS	300–301	NSOPF	135–147	SLS	93–97
B&B	169–177	FRSS	279–283	NAEP	187–205	PEQIS	283–286	SSOCS	291–292
BPS	161–167	HS&B	81–92	NALS	231–242	PLS	99–104	StLA	111–114
CCD	19–25	HST	292–300	NELS:88	53–66	PSS	27–34	TFS	47–51
CivEd	301–304	IALS	243–253	NHES	255–270	SASS	35–46	TIMSS	207–222
CPS	271–277	IEA	223–230	NLS-72	67–79	SCS	287–291		
ECLS	5–18	IPEDS	121–134	NPSAS	149–159	SED	179–186		

E-mail prompts 140, 142

Employees by Assigned Position (EAP) 124

Error checks 240

Estimate 15, 22, 28, 31, 33, 43–44, 75, 99, 103, 108, 113, 143, 145–146, 152, 155, 174–175, 192, 196–197, 199–201, 217–220, 229, 237–239, 248–250, 258–268, 270, 275, 286, 288, 296–297, 303

Estimation methods 2, 13, 22, 24, 30, 41, 49, 60, 74, 88, 95, 101, 108, 113, 118, 130, 141, 154, 164, 174, 182, 196, 217, 227, 237, 249, 262, 275

Ethnographic Case Studies 208, *210*

Excluded Student Survey 187, *189*, 194

Expected Family Contribution (EFC)* 151*, 155

Expected values 13

Expository prose* (i.e., "types of text") 223–224*, 233—see also Prose literacy

F

Faculty Survey 128, *135*, 139–140, 142, 144, 146—see also *Salaries, Tenure, and Fringe Benefits of Full-time Instructional Faculty (SA)*

Fall Enrollment (EF) 123

Fall Enrollment in Occupationally-specific Programs (EP) 124–125, 128

Fall Staff (S) 124

Father Questionnaire 7, 56

Fax follow up 280

Federal Libraries and Information Centers Survey 115–116, 118

FEDLINK* 116*–117

Field edit 236, 240

Field of Doctorate* 180*

Field representatives 41

Field test 11–13, 15, 75, 79, 143–145, 147, 153, 159, 164, 166–167, 173, 176–177, 215, 235, 240, 259, 266–267, 270

Fields of study* 132*

Finance (F) 123

Finance survey—see *School District Finance Survey* and *National Public Education Financial Survey*

First grade 12, 17, 31

First postsecondary institution* 155, 171*

First-stage sampling 70, 220

First-time beginning students (FTBs)* 154, 162*–163

Fiscal data, state-level 22

Flag 22, 60, 130, 260, 264, 299

Focus group 15, 50, 124, 184–185

Followback Study of Excluded Students (FSES) 55—see also *Base Year Ineligible Study*

Follow-up Survey 54, 68, 82, 162, 170—see also *Teacher Follow-up Survey*

Follow-up survey 2, 35, 37, 45, 47–51, 53–55, 57–60, 62–63, 67–69, 71–79, 81–83, 85–86, 88–89, 91, 162, 169–170, 172, 177, 251, 269, 294

Frame 9–10, 16, 27–33, 39, 41–44, 46, 48–50, 57, 70–71, 75–76, 84, 94–95, 100, 118, 122, 126, 142–143, 151–152, 154, 156, 174, 179, 185, 193, 199, 201, 212–213, 225, 229, 234–235, 239, 246, 251, 266, 279–281, 283, 291, 297, 302

Frequency 36, 42, 96, 142, 164, 232, 262, 266, 268, 291–292

Frequency distributions 262

Freshening 9, 57–58, 65

Future plans 2, 15, 23, 31, 42, 50, 67, 82, 96, 102, 110, 113, 118, 131, 142, 155, 165–166, 175, 183, 200, 208, 219, 229, 239, 250, 264, 275, 303

G

Gain scores 7, 64

Gate count* 105, 107*, 116

Generalized Variance Functions (GVFs) 43, 275, 288

Graduation Rate Survey (GRS) 123, 125, 127–128, 130, 134

H

High School Effectiveness Study (HSES) 53–55, 58–59, 66

High School Transcript Study 83, 189–190, *293–294*

High school transcript study 83, 85, 187, 189, 292–293, 297, 300

Home visit 7, 13, 17, 261

Hot-deck 13–14, 31, 41, 50, 130, 142, 155, 228, 238, 264, 275, 281, 284, 296

House weights 218

Household* 3, 6, 8, 14, 16–17, 33, 56, 70, 84, 102, 139, 164, 173, 192, 231–232, 234–240, 244–245, 247, 251, 255–270, 271, 272*–274, 276, 282, 287–290

Household enumeration 259

Household Library Use 257

Household member* 6, 235–236, 258*–261, 263, 266, 271–274, 287–288

Household sample 231–232, 234–235, 237, 239–240

Household survey 3, 24, 33, 245, 262, 265–267, 287

I

Image processing 217—see also Optical scanning

Imaging 41

ALS	105–110	Fed. Lib.	115–119	LCS	300–301	NSOPF	135–147	SLS	93–97
B&B	169–177	FRSS	279–283	NAEP	187–205	PEQIS	283–286	SSOCS	291–292
BPS	161–167	HS&B	81–92	NALS	231–242	PLS	99–104	StLA	111–114
CCD	19–25	HST	292–300	NELS:88	53–66	PSS	27–34	TFS	47–51
CivEd	301–304	IALS	243–253	NHES	255–270	SASS	35–46	TIMSS	207–222
CPS	271–277	IEA	223–230	NLS-72	67–79	SCS	287–291		
ECLS	5–18	IPEDS	121–134	NPSAS	149–159	SED	179–186		

Imputation 13–15, 22, 30–32, 41, 49–50, 60–61, 75, 88, 90, 95, 101, 108, 113, 129–132, 140–142, 147, 154–155, 164–165, 174–175, 198–199, 204, 217, 219–220, 222, 228, 237–238, 249–250, 262–264, 266–267, 269, 274–276, 280–282, 284, 288, 292, 296, 303

Imputation error 220

Incentives 48, 72, 139, 235, 239, 247, 252, 302

Inclusion probabilities 74

Indian school 39—see also BIA school, Private school, and Public school

Institution control* 150, 162*

Institution of Higher Education (IHE)* 126*

Institution sample 138, 151–152, 154–155, 174

Institution Survey 135, 139–146

Institution type* 150*, 155–156, 162*

Institutional Characteristics (IC) 122–123, 127, 151, 283

Instructional faculty* 124–125, 128, 131, 133, 137*–138, 146

Instructional faculty/staff* 137*

Interlibrary loan* 105–106*, 109, 116, 301

International study 219, 222, 230

Internet* 273*—see Web-based survey

Internet reporting option 42

Intersurvey consistency 32, 34

Intraclass correlation 15, 196, 212

IRT scale scores* 7*

Item nonresponse 2, 23, 30, 32, 44, 50–51, 61, 63, 74, 77, 90, 96, 103, 110, 113–114, 118, 132, 141–144, 154, 157, 165–166, 175–176, 184, 201, 229–230, 240, 252, 262, 264, 266, 274, 276, 280–281, 284–285, 289, 292, 297–298

Item response theory (IRT)—see Scaling

Item wording 156–157

Itinerant teacher* 38, 48

J

Jackknife method 15, 43, 156, 165, 200, 220, 229, 237, 239, 251, 281, 284, 288, 292, 297

K

K-terminal school—see Kindergarten-terminal school

Key item—see Critical item

Keyfitz approach 283

Kindergarten Cohort Study 5, 7

Kindergarten-terminal school 28–30, 32–33

L

Leavers* 47, 48*, 49, 85–86
Level of school* 27, 179, 273*
Librarian* 37–38, 42, 93, 94*–96, 105–107, 109, 117
Library/Information Center* 115–116*
Library media center* 31, 36–38*, 40, 42, 93, 94*–96, 112, 117
Library Media Center Survey 37, 40, 93
Library media specialist* 37, 93, 94*–95
Library Media Specialist/Librarian Survey 37, 94
Library survey 2, 37, 93–97, 109–110, 115
Linking study 207–208, 210
List-assisted method 258, 266
List frame 28–33, 39, 48, 95, 201
Literacy* 3, 10, 94, 100, 187–188, 203, 211, 213, 215, 218, 223–233*, 234–244*, 245–253, 255–256, 268, 301
Literacy assessment 227, 231–233, 235, 237–238, 240–241, 244
Literacy Assessment 244
Literacy scales* 224, 232–233*, 235, 238–239, 241, 243, 245*, 249–251
Local Education Agency (LEA)* 35, 38*–39, 94
Logic edits 262
Logical imputation 142, 155
Longitudinal edits 275
Longitudinal sample 5, 53, 59, 67, 81, 152–153, 161, 169
Longitudinal study 2–3, 5, 15, 24, 53, 65–67, 69, 78–79, 81, 83, 91–92, 149, 152–153, 155, 161, 167, 169, 177, 205, 268, 286, 292, 296, 300

M

Machine editing 60, 73, 87, 227, 280, 284
Mail survey 86, 183, 293
Mailback 29–30, 40, 139, 179
Mailout 29–30, 40, 127–130, 139, 154, 291
Makeup session 59, 198, 302
Manual imputation—see Clerical imputation
Marginal maximum likelihood (MML) estimation 197, 199, 201, 249
Matrix sampling 195, 235, 237, 246
Mean square error 43, 239
Measurement error 2, 17, 23, 32, 44–46, 51, 63–64, 66, 77, 90, 92, 96, 103, 110, 114, 118, 132, 144, 147, 157, 159, 165–167, 176–177, 184–185, 199, 201, 222, 229, 240, 252, 266–267, 270, 276, 282, 284, 289, 298

ALS	105–110	Fed. Lib.	115–119	LCS	300–301	NSOPF	135–147	SLS	93–97
B&B	169–177	FRSS	279–283	NAEP	187–205	PEQIS	283–286	SSOCS	291–292
BPS	161–167	HS&B	81–92	NALS	231–242	PLS	99–104	StLA	111–114
CCD	19–25	HST	292–300	NELS:88	53–66	PSS	27–34	TFS	47–51
CivEd	301–304	IALS	243–253	NHES	255–270	SASS	35–46	TIMSS	207–222
CPS	271–277	IEA	223–230	NLS-72	67–79	SCS	287–291		
ECLS	5–18	IPEDS	121–134	NPSAS	149–159	SED	179–186		

Methodological studies 32

Metropolitan statistical area 9, 55, 58, 94, 213, 249

Mitofsky-Waksberg method 258–258, 262

Modal grade* 65, 198, 273*

Movers* 47, 48*–50

Multiple regression—see Regression

Multistage sampling 192, 225

N

NAEP/TIMSS-R Linking Study 208, *210*

Narrative prose* 223–224*, 233—see also Expository prose

National Adult Literacy Survey 3, 203, 231–232, 238, 241–243, 246, 249–250, 252, 268

National-level Assessments 188

National Public Education Financial Survey (NPEFS) 19–*20*, 23

National Questionnaire 224

National Research Coordinators (NRCs)* 181–182, 211*, 214–217, 220–221, 223

Nationally desired population* 211*

Network and Cooperative* 116*

Newborns 10

Nonfiscal data, state-level 22

Noninstructional faculty* 135, 137*–138, 140

Nonresponse bias 12, 16–17, 62–63, 89–90, 154, 165, 239, 251, 262–264

Nonresponse error 2, 16, 23, 32, 44, 50, 61, 75–76, 89, 96, 102, 110, 114, 131, 144, 156, 165, 175, 183, 201, 220, 229, 239, 251, 266, 278, 283, 286

Nonsampling error 2, 15, 23, 31, 44, 50, 61, 75, 89, 96, 102, 110, 114, 118, 131, 143, 156, 165, 175, 183, 200–201, 220, 229, 239, 251, 265, 275–276, 284, 297

Nontraditional students* 162*

Nursery school* 268, 273*, 276

O

Occupationally-specific program* 126*

October Supplement 271–272, 275–276

Online Public Access Catalogue (OPAC)* 107*

OPEID code* 126*

Open-ended 73, 87, 140–141, 199, 205, 227, 236, 241, 247–248, 251, 262

Optical scanning 59, 73, 86–87, 195, 217

Oral Reading Study Assessment 190

Ordinary least squares (OLS) 155

ALS	105–110	Fed. Lib.	115–119	LCS	300–301	NSOPF	135–147	SLS	93–97
B&B	169–177	FRSS	279–283	NAEP	187–205	PEQIS	283–286	SSOCS	291–292
BPS	161–167	HS&B	81–92	NALS	231–242	PLS	99–104	StLA	111–114
CCD	19–25	HST	292–300	NELS:88	53–66	PSS	27–34	TFS	47–51
CivEd	301–304	IALS	243–253	NHES	255–270	SASS	35–46	TIMSS	207–222
CPS	271–277	IEA	223–230	NLS-72	67–79	SCS	287–291		
ECLS	5–18	IPEDS	121–134	NPSAS	149–159	SED	179–186		

Out-of-range 13, 30, 60, 87, 101, 113, 140

Oversampling 9, 40, 60, 65, 70–71, 138–139, 192–194, 234, 259, 262, 269, 292, 298

P

Paper forms 121, 123, 129, 301

Parameter 60, 64, 131, 165, 174, 219–220, 228, 238, 245, 250, 288

Parent and Family Involvement in Education 256–*257*, 264

Parent Interview 6, 8, 12, 14, 16–17, 149–*150*, 152, 157, 170, 257

Parent Questionnaire 14, 16, *54*, 56, 58–59, *82*, 86, 153

Parent/guardian Interview 6

Performance assessment 204, 208–*209*, 211, 213, 216, 219

Periodicity 1, 7, 20, 27, 37, 48, 55, 69, 83, 94, 99, 106, 111, 115, 125, 136, 150, 162, 170, 180, 190, 210, 224, 232, 244, 257, 272

Persistence* 55, 82, 150, 161–162*, 169–170, 285

Pilot survey 246, 252

Plausible value 199, 201, 217, 219–220, 237–239, 249–250

Population—see Target population

Population of the Legal Service Area* 100*, 102–103

Postsecondary education* 3, 53–56, 59, 67–69, 72–75, 78, 82–83, 85, 88–89, 91, 105–107, 112, 121–126*, 128–129, 131–132, 134, 137–138, 149–151, 153–154, 156, 161–163, 174, 279–280, 283–286

Postsecondary Education Transcript Study (PETS) 67, *69*, 72–73, 75, 78, *83*

Poststratification adjustment 146, 154

Precision 43, 138, 146, 151, 155, 165, 201, 212 239, 259, 261

Precoded 262

Pretest 117, 235, 301

Primary sampling units (PSUs) 9–10, 14, 29–31, 33, 39, 191–192, 197, 213, 220, 225–226, 228, 234–237, 239, 246, 249, 273, 287, 293, 302

Principal Survey—see *School Administrator Survey and School Principal Survey*

Prison sample 231–232, 234–235, 237, 239–240

Private school* 2, 8–9, 27, 28*–30, 32–35, 37, 38*–39, 41–42, 44–48, 63, 93, 95, 192, 201, 211, 268, 271–273, 280, 302—see also BIA school, Charter school, and Public school

Private School Universe Survey 2, *27*, 33–34, 280

Probability proportionate to size (pps) 212

Probability sample 39, 53, 55, 57–58, 65, 70, 84–85, 143, 209, 212–213, 245, 252, 272–274, 287, 294, 299

Proc Impute 142

Processing—see Data processing

Proficiency distribution 197

Proficiency probability scores* 8*

Proficiency values 219, 250

Prose 223–224, 231–233, 235, 238, 241, 243–247, 249, 252

Prose literacy* 231, 233*, 244*, 246

Psychometric 10, 12, 63, 66, 79, 92, 246, 250

Psychomotor development 10

Public Charter School Questionnaire 36

Public Education Agency* 19–21*, 25

Public Education Agency Universe Survey 19, *20*

Public elementary/secondary school* 21*, 25

Public Libraries Survey 2, *99*, 114

Public library* 2, 99, 100*–104, 112, 114, 257

Public library service outlet* 100*

Public or private school* 38, 273*

Public school* 9, 19, 21–22, 24–25, 35–38*, 39, 43–46, 48, 53, 63, 93–95, 213, 273, 279–283, 291—see also BIA school, Charter school, Private school

Public School Universe Survey 19

Publication criteria 33

Purpose 1, 5, 12, 15, 19, 27, 29, 35, 39–40, 47, 51, 53, 67, 74, 81, 93, 99, 105, 107, 111, 115–116, 122–126, 135, 137, 145–146, 149, 152, 154, 161, 165–166, 169, 179, 187, 206, 211, 214–215, 218, 221, 229, 241–242, 253–254, 261, 271, 287, 296\

Q

QED 39, 192, 200, 213, 225, 228–229, 302

Quality control 12, 60, 75, 77, 79, 130, 132, 140, 143, 154, 164, 176, 195, 200, 220, 229, 236, 240, 249–251, 303

Quantitative literacy* 232–233*, 241, 244*–246, 249

Questionnaire changes 2, 32–33, 133, 143

Questionnaire printing 42

R

Race/ethnicity* 2, 6, 8*–10, 16, 20, 23–24, 27, 33, 61, 75, 90–91, 123–126, 130, 133, 150, 155, 157, 161, 165, 170, 179, 182, 184–185, 189, 219, 232, 237–238, 246, 263, 268, 294, 296–298

Raking 237, 263, 266, 274

Random digit dialing (RDD) 258–259, 280

Range check 30, 41, 49, 73, 95, 107–108, 154, 227, 248, 262

Ratio adjustment 41, 48, 75, 108, 237, 249, 276

Reading assessment—see Literacy assessment

Reading Literacy Tests 223, 227–228

ALS	105–110	Fed. Lib.	115–119	LCS	300–301	NSOPF	135–147	SLS	93–97
B&B	169–177	FRSS	279–283	NAEP	187–205	PEQIS	283–286	SSOCS	291–292
BPS	161–167	HS&B	81–92	NALS	231–242	PLS	99–104	StLA	111–114
CCD	19–25	HST	292–300	NELS:88	53–66	PSS	27–34	TFS	47–51
CivEd	301–304	IALS	243–253	NHES	255–270	SASS	35–46	TIMSS	207–222
CPS	271–277	IEA	223–230	NLS-72	67–79	SCS	287–291		
ECLS	5–18	IPEDS	121–134	NPSAS	149–159	SED	179–186		

Recall 77, 91, 157, 184, 289–290

Recent changes 31–32, 41, 50, 96, 101, 109, 113, 130, 142, 155, 182, 199, 250, 264

Record Studies 81, *83*

Reference dates 11, 21, 29, 40, 49, 58, 72, 86, 95, 101, 107, 112, 117, 127–128, 139, 153, 163, 173, 181, 195, 215, 226, 236, 247, 261, 274

Reference transaction* 99, 101, 103, 105–106*, 300

Refusal 49, 57, 59–60, 76, 89–90, 164, 173, 238, 240, 250, 289, 293

Refusal conversion 173

Regression analysis 203

Regular school 21, 190, 264, 271–273, 300

Reinterview 45, 51, 96–97, 145–146, 154, 158, 166, 176, 264, 266–267, 270

Relational edit check 101, 108, 113, 117

Reliability 17, 24, 63, 90, 96, 110, 154, 158, 166, 192, 196, 199, 221, 240, 264, 266, 273, 276, 295

Reminder postcard 30, 40, 59, 72

Replicate 15, 43, 50, 143, 156, 165, 229, 237, 239, 251, 265, 275, 297

Replicate estimate 43, 239, 297

Replicate weight 5, 43, 50, 143, 156, 165, 229, 237, 251, 265, 297

Replication technique 288

Reporting period 102–103, 114, 133

Reporting period differences 78, 103

Research doctorate* 179–180*

Response bias 15, 62, 252, 270

Response rate 10, 14, 16–17, 23, 32, 44, 46, 50–51, 60, 62, 66, 74, 76-78, 96, 102–103, 110, 114, 117–118, 131–132, 138, 142–144, 156–157, 159, 165-167, 175–177, 182–185, 201–202, 220–221, 229–230, 235, 239–240, 246, 251–252, 262–264, 266–267, 269, 281, 283, 288–289, 295, 297, 302—see also Nonresponse error

Response variance 45, 51, 96–97

Rotation 216, 249, 252, 288

Routing 5, 11, 14, 73, 77, 216

S

Salaries (SA) 124

Salaries, Tenure, and Fringe Benefits of Full-time Instructional Faculty 124

Sample design 2, 8–9, 15, 21, 28, 39, 41, 43, 45, 48–49, 57, 61, 64, 66, 70, 75–79, 84–85, 88–89, 91–92, 94–97, 101, 107, 112, 117, 127, 138, 143, 151, 155, 163, 165, 172, 181, 191, 197, 212–213, 217–218, 225, 227, 229, 234, 237–239, 241, 245, 251, 253, 258, 261–262, 264–266, 273, 279, 283, 287–289, 291–292, 295, 297–299, 302

Sample design changes 77

Sample redefinitions and augmentations 71

Sample replacement 89

Sample survey 5, 33, 35, 41, 45, 48–49, 53, 67, 81, 93, 95, 97, 149, 161, 169, 220, 236, 255, 277

Sampling error 15, 23, 31, 43, 50, 61, 75, 78, 88, 96, 102, 110, 114, 118, 130–131, 143, 156, 165, 175, 183, 200–201, 219, 228–229, 239, 251, 265, 275, 281, 284, 288, 292, 296–297, 303

Sampling frame 16, 27, 32, 39, 42, 44, 57, 70–71, 94, 122, 151–152, 156, 179, 193, 201, 212–213, 225, 229, 234–235, 239, 251, 266, 279–281, 283, 291, 297, 302

Sampling households 258

Sampling unit 9, 13, 71, 191, 212–213, 220, 225, 246, 273, 287, 296

Sampling variance 15, 43, 96, 165, 201, 218, 239, 297

Sampling within households 259

Scaling 14, 60, 64, 197–200, 202–203, 217, 219, 228, 237–238, 249–250, 303

Scanning—see Optical scanning

School Administrator Questionnaire 6, 45, *54*, 93

School Characteristics and Policies Survey 187, *189*, 191, 195, 201

School District Finance Survey 19–*20*, 25

School District Survey (formerly titled the Teacher Demand and Shortage Survey-TDS) *35*, 44, 280

School enrollment* 48, 57, 61, 63, 95, 151, 170–171, 268, 271, 272*–273, 275–276, 285

School Facilities Checklist 6

School library media center* 36–37, 40, 93, 112*—see also Library media center

School Library Media Center Survey 40, *93*

School Library Media Specialist/Librarian Survey 94

School Principal Survey (formerly titled the *School Administrator Survey*) 36

School Questionnaire 6, 12, 29, 35–36, 40, 42, 48, 54, *68*, 72, 74, *82*, 84, 89, 93, 95, *209*, *224*, 226–227, 294, 301–*302*

School Readiness—see *Early Childhood Education and School Readiness*

School Safety and Discipline 35, *257*, 268, 270, 290

School Survey 3, 9, 28–29, *36*, 39, 41, 43–44, 94–95, 213, 281, 287, 291, 302

School type 6, 16, 31, 40, 50, 138, 296, 302

Scoring 63, 191, 195–197, 199, 201, 203, 211, 215, 227, 235–237, 240–241, 247–249, 251, 303

Screener 5, 29, 32, 235–236, 238–240, 255, 257–264, 266–267, 289

SD/LEP Survey 187, *189*, 192–195, 198–199

Secondary school* 21, 25, 28*, 41, 44, 47–48, 56, 66, 68, 74, 91, 112, 115, 117, 187–188, 191, 211, 261, 279, 295, 300—see also Combined school and Elementary school

Self-administered questionnaire 11–12, 58, 139, 280, 284

Self-weighting 71, 213, 287, 302

Senate weights 218

September Supplement 271–*272*, 275

Simple random sampling 220, 229, 262

Skip pattern 87, 182, 262, 264

Socioeconomic scale* 8*
Socioeconomic status (SES)* 8, 13–15, 17, 56*, 63, 65, 70*, 84*, 75, 163*–164, 195, 217, 302
Source of support* 181*–182, 184–185
Special education 5–6, 14, 21, 23, 28, 36, 38–39, 42, 44, 57, 95, 189, 281, 293–294, 299
Special education school 28, 57
Special education student 36, 42, 293–294, 299
Special Education Teacher Questionnaire 6
Special library* 112*
Special population 150
Special Studies 190, 275, 277
Spiraling 195, 235
Standard deviation 65, 228
Standard error 15, 31, 33, 43, 61, 75, 88, 143, 200, 220, 229, 259, 275, 281, 284, 288-289, 292, 297, 303
Standardized scores (T-scores)* 7*, 8, 60
State-level Assessments 188, 190, 200
State library agency* 2, 99, 100*–101, 111–114
State Nonfiscal Survey 19–*20*, 23–24
Status in teacher pipeline* 171*
Stayers* 47, 48*–50
StLA Survey 2, *111*–114
Stratification 9, 39, 71, 88, 94, 151, 165, 172, 192, 212, 214, 225, 228, 237, 273, 276, 279–280, 302
Student Background Questionnaire 206, 228
Student Financial Aid (SFA) 123
Student Financial Aid Records 83
Student Interview *150*, 152, 154, 157–158, 169–170, 173, 289
Student Questionnaire 15, *53*–54, 59–61, 63–64, *67*, 71–72, *82*, 91, 206, 226–227, *301*
Student Record Information Form (SRIF) 68, 72
Student Record(s) Abstract 6, 12, *149*, 169
Student Records Component 37, 42
Student sample 9, 54–55, 57–59, 62, 65, 68, 74, 89, 152, 172, 191, 194, 208, 215, 293–294, 299, 303
Subsampling 57, 61, 75, 85, 138, 153, 172, 212–213, 218, 228, 237
Substitution 84, 89, 201, 238, 263
Succeed consistently* 203, 233*, 245
Summation check 107–108
Supplemental Studies 55
Survey of Earned Doctorates 3, 133, *179*, 186
System* (i.e., library system) 112*

ALS	105–110	Fed. Lib.	115–119	LCS	300–301	NSOPF	135–147	SLS	93–97
B&B	169–177	FRSS	279–283	NAEP	187–205	PEQIS	283–286	SSOCS	291–292
BPS	161–167	HS&B	81–92	NALS	231–242	PLS	99–104	StLA	111–114
CCD	19–25	HST	292–300	NELS:88	53–66	PSS	27–34	TFS	47–51
CivEd	301–304	IALS	243–253	NHES	255–270	SASS	35–46	TIMSS	207–222
CPS	271–277	IEA	223–230	NLS-72	67–79	SCS	287–291		
ECLS	5–18	IPEDS	121–134	NPSAS	149–159	SED	179–186		

System, Federation, or Cooperative Service* 100*

Systematic error 239, 249, 266

Systematic sampling 9, 152

T

T-scores—see Standardized scores

Target population 2, 8, 21, 28, 38, 48, 57, 70, 84, 94, 101, 107, 112, 117, 127, 138, 151, 154, 162–163, 171, 181, 191, 211, 220, 224–225, 234, 245, 258, 265, 273

Taylor series 15, 61, 75, 88, 143, 156, 165, 175, 265, 303

Teacher* 2, 6–7, 12, 14–18, 21, 24, 28*, 31, 33, 35–38*, 40–51, 54–55, 58–59, 63, 69, 71–72, 74, 82, 93–94, 138, 145, 150, 170–171, 187–191, 194–195, 198–201, 208–209, 213–214, 216–219, 223–224, 226–227, 230, 256, 279–283, 290, 292, 294, 301–302

Teacher Comment Checklist 82

Teacher Demand and Shortage (TDS) Survey—see *School District Survey*

Teacher Followup Survey Questionnaire for Continuing Teachers 47

Teacher Followup Survey Questionnaire for Former Teachers 47, 51

Teacher pipeline 171

Teacher Questionnaire 6–7, 14, 16, 35–37, 40, 42, 45–46, 50, *54*, 145, 189–190, 201, *208*–209, 213, 223–*224*, 290, 301–*302*

Teacher Survey 36–37, 43–44, 48–50, 187, *189*, 191, 195, 280–281, 290

Teaching supplement 68–69, 72–74, 76

Technology-Based Assessment (TBA) Project 190

Telephone follow up 29–30, 40, 44, 49, 73, 77, 113, 129, 132, 139–140, 291

Telephone interviewing 30, 40–41, 59, 86, 90, 139–140, 152–153, 155, 163, 173, 261, 274, 288—see also CATI

Test battery* 56, 63–64, 67–69*, 71–72, 82

Time series 185

Time-to-doctorate* 180*, 185

TIMSS-R Benchmarking Project 210

Title* 106*

Title IV financial aid* 127, 130, 151*

Transcript 3, 53–55, 58–59, 65–67, 69, 72–76, 78, 81–88, 92, 169–170, 172–175, 177, 187, 189–190, 204–205, 286–287, 292–300

Transcript Study 54–55, 58, 67, 69, 72–73, 75, 78, 83, 85, 88, 170, 187, 189, 204, 292, *293*, *294*, 296–300—see also *High School Transcript Study* and *Postsecondary Education Transcript Study*

Trend data 53, 55, 83, 215

Two-stage sampling 280

U

Unbiased estimate 297

Undercoverage 31, 39, 44, 61, 71, 75–76, 102, 157, 183, 239, 262, 265–266, 270, 276, 281, 297

Underestimate 240

Unit nonresponse 16, 23, 32, 44, 50, 61–62, 74, 76–77, 88–89, 96, 102, 110, 114, 118, 131, 141, 144, 154, 156, 164–165, 174–175, 183, 201, 220, 229–230, 239, 249, 251, 262, 266, 274, 276, 282, 284, 289, 292, 297

Universe survey 2, 19–20, 22–23, 25, 27, 33–34, 44, 102, 110, 114, 118, 280–281, 290

V

Validity 24, 44, 63–64, 129, 195–197, 239, 251

Variance 15, 17, 42–46, 50–51, 96–97, 143–144, 165, 201, 218, 220, 237–239, 250–251, 264–265, 275, 288, 297

Variance estimator 43, 143

Videotape 207–209, 214, 216–217

Videotape Study 208–209, 214, 216

Vocational education 19, 23, 39, 44, 95, 122, 124, 150, 272, 279, 281–282

Vocational school 21, 57, 69, 77, 273

Vocational school enrollment* 273*

Volume* 43–46, 94, 97, 106*, 210, 222, 301

W

Web-based survey 37, 40, 107, 130

Web collection 107, 110, 125, 131

Web questionnaire 139–140

Weighting 13, 30, 41–42, 49–50, 60, 74–75, 77, 88, 95, 130, 141, 144, 154, 164, 174, 177, 182, 197, 200, 217–219, 228–229, 237–240, 249–250, 262–263, 265–267, 269, 274–275, 280, 284, 287–288, 292, 295, 303

Wording 29, 42, 51, 64–65, 141, 156–157, 214, 221, 227, 285

Written Assessment 208–209, 212–213, 219

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").

EFF-089 (1/2003)